Rasmus Larsen
Mads Nielsen
Jon Sporring (Eds.)

Medical Image Comp and Comput Intervention

LNCS 4190

9th International Conferen
Copenhagen, Denmark, Oc
Proceedings, Part I

Lecture Notes in Computer Science 4190

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

Rasmus Larsen Mads Nielsen
Jon Sporring (Eds.)

Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006

9th International Conference
Copenhagen, Denmark, October 1-6, 2006
Proceedings, Part I

Springer

Volume Editors

Rasmus Larsen
Technical University of Denmark
Informatics and Mathematical Modelling
2800 Kgs. Lyngby, Denmark
E-mail: rl@imm.dtu.dk

Mads Nielsen
IT University of Copenhagen
Rued Langgaards Vej 7, 2300 København S, Denmark
E-mail: malte@itu.dk

Jon Sporring
University of Copenhagen
Department of Computer Science
Universitetsparken 1, 2100 København Ø, Denmark
E-mail: sporring@diku.dk

Library of Congress Control Number: 2006932793

CR Subject Classification (1998): I.5, I.4, I.3.5-8, I.2.9-10, J.3, J.6

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

ISSN 0302-9743
ISBN-10 3-540-44707-5 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-44707-8 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11866565 06/3142 5 4 3 2 1 0

Preface

The 9th International Conference on Medical Image Computing and Computer Assisted Intervention, MICCAI 2006, was held in Copenhagen, Denmark at the Tivoli Concert Hall with satellite workshops and tutorials at the IT University of Copenhagen, October 1-6, 2006.

The conference has become the premier international conference with in-depth full length papers in the multidisciplinary fields of medical image computing, computer-assisted intervention, and medical robotics. The conference brings together clinicians, computer scientists, engineers, physicists, and other researchers and offers a forum for the exchange of ideas in a multidisciplinary setting.

MICCAI papers are of high standard and have a long lifetime. In this volume as well as in the latest journal issues of Medical Image Analysis and IEEE Transactions on Medical Imaging papers cite previous MICCAIs including the first MICCAI conference in Cambridge, Massachusetts, 1998. It is obvious that the community requires the MICCAI papers as archive material. Therefore the proceedings of MICCAI are from 2005 and henceforth being indexed by Medline.

A careful review and selection process was executed in order to secure the best possible program for the MICCAI 2006 conference. We received 578 scientific papers from which 39 papers were selected for the oral program and 193 papers for the poster program.

The papers were evaluated by 3 independent scientific reviewers. Reviewer affiliations were carefully checked against author affiliations to avoid conflicts of interest, and the review process was run as a double blind process. A special procedure was devised for papers from the universities of the organizers upholding a double blind review process also for these papers. A total of 98 % of the reviews we asked for were received.

The MICCAI program committee consisted of the local organizers, 2 internationally selected co-chairs, and 15 internationally selected area chairs, each a leading expert in his/her field. Each area chair was assigned 40 papers from which he/she formed a recommendation for the program committee based on the scientific reviews as well as their own assessment.

The entire program committee met in Copenhagen for 2 full days in May 2006. At this meeting all 578 papers and their corresponding reviews were printed and discussed. In a first round of discussions the area chairs were divided into 5 groups of 3. From their joint pool of 120 papers each group identified 12 potential oral papers, 28 poster papers, and 16 potential poster papers. In the second round the oral program was made from the resulting 60 potential oral papers. Oral papers were selected based on their quality, total coverage of MICCAI topics, and suitability for oral presentation. In parallel the remaining 80 potential poster papers were considered and 33 papers were accepted for poster presentations.

The entire procedure was designed such that the papers were selected by paper to paper comparison forcing the committee members to argue for the decision in each individual case.

We believe a careful and fair selection process has been carried out for MICCAI 2006. Each paper was examined by 3 reviewers, and further scrutinized by 3-8 program committee members. Our thanks go to the reviewers and area chairs for their hard work and enthusiasm, and to the two program co-chairs David Hawkes and Wiro Niessen for their dedication to putting together the program.

This year's MICCAI was augmented by more workshops than previously. Twelve independent workshops were held prior and subsequent to the conference. These workshops served as a forum for MICCAI subfields and made room for many more presentations due to their parallel programs. The workshops were organized for all scientific and most practical matters by the workshop chairs. We thank the workshop organizers for suggesting, arranging, and managing these excellent workshops. It is our hope that we will see multiple workshops also at future MICCAI conferences. Three tutorials were also provided by leading experts in their fields of research.

We thank the two keynote speakers Terry Jernigan, UCSD and Copenhagen University Hospital, Hvidovre, Denmark and Thomas Sinkjær, Director, Center for Sensory-Motor Interaction, Aalborg University, Denmark. A series of sponsors helped make the conference possible. for this they are thanked. Finally, we thank our 3 general co-chairs Anthony Maeder, Nobuhiko Hata, and Olaf Paulson, who provided insightful comments and invaluable support during the entire process of planning MICCAI 2006.

The greater Copenhagen region in Denmark and the Skåne region in Southern Sweden are connected by the Øresund Bridge. The region hosts 14 universities and a large concentration of pharmaceutical and biotech industry as well as 26 hospitals. This makes Copenhagen the capital of one of the most important life science centers in Europe.

It was our great pleasure to welcome delegates from all over the world to Denmark and the city of Copenhagen. It is our hope that delegates in addition to attending the conference took the opportunity to sample the many excellent cultural offerings of Copenhagen.

We look forward to welcoming you to MICCAI 2007 to be held October 29 - November 2 in Brisbane, Australia and chaired by Anthony Maeder.

October 2006 Rasmus Larsen, Mads Nielsen, and Jon Sporring

Organization

The university sponsors for MICCAI 2006 were the IT-University of Copenhagen, the Technical University of Denmark, and the University of Copenhagen.

Executive Committee

General Chairmanship

Mads Nielsen (chair)	IT-University of Copenhagen, Denmark
Nubohiko Hata	Brigham and Women's Hospital, Boston, USA
Anthony Maeder	University of Queensland, Brisbane, Australia
Olaf Paulson	Copenhagen University Hospital, Denmark

Program Chairmanship

Rasmus Larsen (chair)	Technical University of Denmark
David Hawkes	University College London, UK
Wiro Niessen	Erasmus Medical School, Netherlands

Workshop and Tutorials Chair

Jon Sporring	University of Copenhagen, Denmark

Program Committee

Leon Axel	New York University Medical Center, USA
Marleen de Bruijne	IT University of Copenhagen, Denmark
Kevin Cleary	Georgetown University Medical Center, USA
Hervé Delingette	INRIA, Sophia Antipolis, France
Polina Golland	Massachusetts Institute of Technology, USA
Nico Karssemeijer	Radboud University Nijmegen, Netherlands
Sven Kreiborg	University of Copenhagen, Denmark
Jyrki Lötjönen	VTT, Finland
Kensaku Mori	Nagoya University, Japan
Sébastien Ourselin	CSIRO, Australia
Egill Rostrup	University of Copenhagen, Denmark
Julia Schnabel	University College London, UK
Pengcheng Shi	Hong Kong University of Science and Technology
Martin Styner	University of North Carolina, USA
Carl-Fredrik Westin	Harvard University, USA

Conference Secretariat and Management

Camilla Jørgensen	IT University of Copenhagen, Denmark
Eva Branner	Congress Consultants, Denmark
Henrik J. Nielsen	Congress Consultants, Denmark

Student Awards Coordinator

Karl Heinz Höhne	Germany

Local Organizing Committee

Erik Dam	IT University of Copenhagen
Sune Darkner	Technical University of Denmark
Søren Erbou	Technical University of Denmark
Mads Fogtmann Hansen	Technical University of Denmark
Michael Sass Hansen	Technical University of Denmark
Peter Stanley Jørgensen	Technical University of Denmark
Marco Loog	IT University of Copenhagen
Hildur Ólafsdóttir	Technical University of Denmark
Ole Fogh Olsen	IT University of Copenhagen
Mikkel B. Stegmann	Technical University of Denmark
Martin Vester-Christensen	Technical University of Denmark

Sponsors

AstraZeneca
Center for Clinical and Basic Research
Claron
Elsevier
GE
Medtronic
NDI - Northern Digital Inc.
Siemens Corporate Research
Springer
Visiopharm

Reviewers

Hossam El Din Hassan Abd El Munim
Purang Abolmaesumi
Elsa Angelini
Anastassia Angelopoulou
Neculai Archip
John Ashburner
Stephen Aylward
Fred S. Azar
Jose M. Azorin
Eric Bardinet
Christian Barillot
Philip Batchelor
Pierre-Louis Bazin
Fernando Bello
Marie-Odile Berger
Abhir Bhalerao
Rahul Bhotika
Isabelle Bloch
Emad Boctor
Thomas Boettger
Hrvoje Bogunovic
Sylvain Bouix
Pierrick Bourgeat
Roger Boyle
Elizabeth Bullitt
Catherina R. Burghart
Darius Burschka
Nathan Cahill
Hongmin Cai
Darwin G. Caldwell
Oscar Camara-Rey
Carlos Alberto Castao Moraga
Pascal Cathier
M. Mallar Chakravarty
Hsun-Hsien Chang
Jian Chen
Lishui Cheng
Aichi Chien
Kiyoyuki Chinzei
Gary Christensen
Albert C.S. Chung
Moo Chung
Chris A. Cocosco
D. Louis Collins
Olivier Colliot
Lars Conrad-Hansen
Jason Corso
Olivier Coulon
Patrick Courtney
Jessica Crouch
Erik Dam
Mikhail Danilouchkine
Sune Darkner
Julien Dauguet
Laura Dempere-Marco
Maxime Descoteaux
Michel Desvignes
Maneesh Dewan
Jean-Louis Dillenseger
Simon DiMaio
Christophe Doignon
Etienne Dombre
Andrew Dowsey
Ye Duan
Simon Duchesne
Ayman El-Baz
Randy Ellis
Søren Erbou
Simon Eskildsen
Yong Fan
Aly Farag
Aaron Fenster
Gabor Fichtinger
Oliver Fleig
P. Thomas Fletcher
Charles Florin
Mads Fogtmann Hansen
Jenny Folkesson
Rui Gan
Andrew Gee
Guido Gerig
David Gering
Frans Gerritsen
Bernard Gibaud
Maryellen Giger
Gaolang Gong

Ren Hui Gong
Miguel Á.G. Ballester
Mark Gooding
Girish Gopalakrishnan
Vicente Grau
Eric Grimson
Christophe Grova
Christoph Guetter
Efstathios Hadjidemetriou
Horst Hahn
Haissam Haidar
Ghassan Hamarneh
Lars G. Hanson
Matthias Harders
Makoto Hashizume
M. Sabry Hassouna
Mark Hastenteufel
Peter Hastreiter
Yong He
Pierre Hellier
David Holmes
Byung-Woo Hong
Robert Howe
Qingmao Hu
Zhenghui Hu
Heng Huang
Karl Heinz Höhne
Ameet Jain
Pierre Jannin
Branislav Jaramaz
Tianzi Jiang
Yuchong Jiang
Ge Jin
Leigh Johnston
Julien Jomier
Sarang Joshi
Leo Joskowicz
Ioannis Kakadiaris
D.B. Karron
Michael Kaus
Peter Kazanzides
Kamran Kazemi
Erwan Kerrien
Irina Boshko Kezele
Ali Khamene
Ron Kikinis
Adelaide Kissi
Takayuki Kitasaka
Jan Klein
Ender Konukoglu
Tim Kroeger
Thomas Lange
Thomas Lango
Rudy Lapeer
Sang-Chul Lee
Koen van Leemput
Chunming Li
Shuo Li
Jianming Liang
Hongen Liao
Rui Liao
Yuan-Lin Liao
Jean Lienard
Marius George Linguraru
Alan Liu
Huafeng Liu
Tianming Liu
Marco Loog
William Lorensen
Peter Lorenzen
Anant Madabhushi
Mahnaz Maddah
Frederik Maes
Sherif Makram-Ebeid
Gregoire Malandain
Robert Marti
Marcos Martin-Fernandez
Ken Masamune
Julian Mattes
Tim McInerney
Gloria Menegaz
Chuck Meyer
Michael I. Miga
James Miller
Abhilash Miranda
Lopamudra Mukherjee
William Mullally
Yoshihiro Muragaki
Delphine Nain
Kyojiro Nambu

Sumit Nath
Nassir Navab
Stephane Nicolau
Marc Niethammer
Alison Noble
Herke Jan Noordmans
Wieslaw L. Nowinski
Thomas O'Donnell
Arnaud Ogier
Allison M. Okamura
Silvia Olabarriaga
Hildur Ólafsdóttir
Salvador Olmos
Ole Fogh Olsen
Mark Olszewski
Tobias Ortmaier
Xenophon Papademetris
Nikos Paragios
Hyunjin Park
Javier Pascau
Rajni Patel
Alexandru Patriciu
Perrine Paul
Rasmus Paulsen
Ioannis Pavlidis
Kim Steenstrup Pedersen
Heinz-Otto Peitgen
Mélanie Pelegrini-Issac
Xavier Pennec
Dimitrios Perperidis
Eric Pichon
Josien Pluim
Kilian Pohl
Richard Prager
Tobias Preusser
Sylvain Prima
Jerry L. Prince
Yingge Qu
Srinivasan Rajagopalan
Nasir Rajpoot
Richard A. Robb
Miguel Angel Rodriguez-Florido
Torsten Rohlfing
Karl Rohr
Michael Rosenthal
Daniel Rueckert
Daniel Russakoff
Ichiro Sakuma
Tim Salcudean
Thomas Sebastian
Zuyao Shan
Cartik Sharma
Dinggang Shen
Hongjian Shi
Lin Shi
Rudolf Sidler
Alberto Signoroni
Nabil Simaan
Vikas Singh
Karl Sjöstrand
Örjan Smedby
Xubo Song
Jon Sporring
James Stewart
Rik Stokking
Danail Stoyanov
Yi Su
Navneeth Subramanian
Paul Suetens
Gábor Székely
Songyuan Tang
Xiaodong Tao
Huseyin Tek
Demetri Terzopoulos
Jean-Philippe Thiran
Marc Thiriet
Carlos Thomaz
Jussi Tohka
Oliver Tonet
Shan Tong
Jocelyne Troccaz
Gözde Ünal
Regis Vaillant
Ragini Verma
Martin Vester-Christensen
Pierre Vieyres
Kirby Vosburgh
Albert Vossepoel
Lionel C. C. Wai
Defeng Wang

Linwei Wang
Qiang Wang
Yiying Wang
Yongmei Michelle Wang
Yuanquan Wang
Simon Warfield
Zhouping Wei
Ross Whitaker
James Williams
Cees van Wijk
Ivo Wolf
Wilbur C.K. Wong
Chia-Hsiang Wu
John Jue Wu
Ting Wu
Chris Wyatt
Stefan Wörz
Zhong Xue
Yasushi Yamauchi
Pingkun Yan
G.Z. Yang
Ziv Yaniv
Terry Yoo
Paul Yushkevich
Stefan Zachow
Jianchao Zeng
Yiqiang Zhan
Zheen Zhao
Guoyan Zheng
S. Kevin Zhou
Wanlin Zhu
Tatjana Zrimec
Reyer Zwiggelaar

MICCAI Society

Executive Officers

Alan Colchester	President and Board Chair
Richard A. Robb	Executive Director
Nicholas Ayache	Executive Secretary
Terry M. Peters	Treasurer
Karl Heinz Höhne	Elections Officer (Honorary Board member)

Staff

Gábor Székely	Membership Coordinator
Nobuhiko Hata	Publication Coordinator

Board of Directors

Nicholas Ayache	INRIA, Sophia Antipolis, France
Alan Colchester	University of Kent, Canterbury, UK
James Duncan	Yale University, New Haven, Connecticut, USA
Guido Gerig	University of North Carolina, Chapel Hill, USA
Anthony Maeder	University of Queensland, Brisbane, Australia
Dimitris Metaxas	Rutgers University, New Jersey, USA
Mads Nielsen	IT University of Copenhagen, Copenhagen, Denmark
Alison Noble	University of Oxford, Oxford, UK
Terry M. Peters	Robarts Research Institute, London, Ontario, Canada
Richard A. Robb	Mayo Clinic College of Medicine, Rochester, Minnesota, USA

Student Awards

Every year MICCAI awards outstanding work written and presented by students. Both oral and poster presentations are eligible for the awards, and the awards are presented to the winners in a public ceremony at the conference.

MICCAI 2005 – Palm Springs

At MICCAI 2005 five prizes each valued at 500 USD sponsored by Northern Digital Incorporation (NDI) were awarded in the following categories

Image segmentation and analysis: Pingkun Yan, MRA Image Segmentation with Capillary Active Contour

Image registration: Ashraf Mohamed, Deformable Registration of Brain Tumor Images via a Statistical Model of Tumor Induced Deformation

Computer assisted interventions and robotics: Henry C. Lin, Automatic Detection and Segmentation of Robot Assisted Surgical Motions

Simulation and visualization: Peter Savadjiev, 3D Curve Inference for Diffusion MRI Regularization

Clinical applications: Srinivasan Rajagopalan, Schwarz Meets Schwann: Design and Fabrication of Biomorphic Tissue Engineering Scafolds

MICCAI 2004 – St. Malo

At MICCAI 2004 four prizes each valued at 600 Euros sponsored by Northern Digital Incorporation (NDI) were awarded in the following categories

Image segmentation and processing: Engin Dikici, Quantification of Delayed Enhancement MR Images

Image registration and analysis: Dimitrios Perperidis, Spatio-Temporal Free-Form Registration of Cardiac MR Image Sequences

Image guided therapy and robotics: Danail Stoyanov, Dense 3D Depth Recovery for Soft Tissue Deformation During Robotically Assisted Laparoscopic Surgery

Image Simulation and Display: Davide Valtorta, Dynamic Measurements of Soft Tissue Viscoelastic Properties with a Torsional Resonator Device

Table of Contents – Part I

Bone Shape Analysis

Robotics and Tracking

Segmentation

Analysis of Diffusion Tensor MRI

Shape Analysis and Morphometry

Simulation and Interaction

Robotics and Intervention

Cardio-vascular Applications

Image Analysis in Oncology

Brain Atlases and Segmentation

Cardiac Motion Analysis

Clinical Applications I

Registration I

Table of Contents – Part II

Segmentation I

Validation and Quantitative Image Analysis

Brain Image Processing

Motion in Image Formation

Image Guided Intervention

Clinical Applications II

Registration II

Segmentation II

Brain Analysis and Registration

Quantitative Vertebral Morphometry Using Neighbor-Conditional Shape Models

Marleen de Bruijne[1], Michael T. Lund[1],
László B. Tankó[2], Paola P. Pettersen[2], and Mads Nielsen[1]

[1] IT University of Copenhagen, Denmark
[2] Center for Clinical and Basic Research, Ballerup, Denmark
marleen@itu.dk

Abstract. A novel method for vertebral fracture quantification from X-ray images is presented. Using pairwise conditional shape models trained on a set of healthy spines, the most likely normal vertebra shapes are estimated conditional on all other vertebrae in the image. The differences between the true shape and the reconstructed normal shape is subsequently used as a measure of abnormality. In contrast with the current (semi-)quantitative grading strategies this method takes the full shape into account, it uses a patient-specific reference by combining population-based information on biological variation in vertebra shape and vertebra interrelations, and it provides a continuous measure of deformity.

The method is demonstrated on 212 lateral spine radiographs with in total 78 fractures. The distance between prediction and true shape is 1.0 mm for unfractured vertebrae and 3.7 mm for fractures, which makes it possible to diagnose and assess the severity of a fracture.

1 Introduction

Osteoporosis is a common skeletal disorder characterized by a decrease in bone mass, leading to bone fragility and an increased risk of fractures. It is a major public health problem; one out of every three women and one out of eight men over the age of 50 is expected to have an osteoporosis-related fracture in the remainder of their lives. Any bone can be affected but the fractures typically occur in the hip, spine, and wrist. Of these, hip fractures are the most serious in terms of morbidity and mortality. Vertebral fractures can be asymptomatic, but can also have serious consequences, including severe back pain and deformity. Furthermore, vertebral fractures are the most common osteoporotic fracture, they occur in younger patients, and their presence is known to be a good indicator for the risk of future spine and hip fractures. This makes vertebral fracture assessment suitable as an outcome variable in clinical trials to diagnose osteoporosis incidence and progression.

Vertebral fractures are conventionally detected and graded on lateral X-rays. Numerous methods have been proposed for this purpose, typically relying on a subjective judgement of vertebral appearance by an expert radiologist in connection with six-point morphometry. In the latter, six points are placed on the corners and in the middle of the vertebra endplates, defining the anterior, middle and posterior heights. The fracture

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 1–8, 2006.

grade is derived from these heights or from the ratios between these heights, possibly in comparison with population based measurements and/or normalized for inter-patient variability by comparison with measurements taken from a neighboring or reference vertebra (see for instance [1, 2, 3, 4]).

Several authors have proposed methods for automatic spine segmentation from X-ray or dual X-ray absorptiometry (DXA) images with the aim of automating vertebral morphometry [5, 6, 7, 8]. In this work, we start from the segmented vertebrae and aim at improving the diagnosis of fractures with respect to the current height based measurements. Currently, the diagnosis of a vertebral fracture in clinical trials relies on a 20% decrease in body height of a vertebra. Due to the categorical nature of this study parameter, clinical trials continue to demand a large number of participants and long follow-up time.

One shortcoming of the conventional models is that the sparse representation of six points is unable to capture subtle shape changes. Smyth et al. [5] used point distribution models to represent the full contour of normal and fractured vertebrae and detected fractures using a Mahalanobis distance classifier on this representation. A slight, but significant improvement with respect to conventional height measurements was reported.

We propose to model not only the shape variation over a population for individual vertebrae, but to also model the interrelations between vertebrae in the same subject. This additional prior information allows adjustment of the models to individual patients so as to distinguish normal biological shape variation from osteoporosis-related deformation. We use conditional shape models to predict the most likely shape of a vertebra given the known shape of a neighboring vertebra. If the models are constructed from a training set of normal, healthy spines this provides an estimate of what the vertebra shape would have been if it were normal. In a previous paper we showed that vertebrae in healthy spines can be accurately reconstructed from their neighbors[9].

In the current paper we show how the reliability of predicted shapes can be estimated and we use this to combine pairwise predictions of all vertebrae in the image into one single, optimal prediction for each vertebra. The difference between the prediction and the true, segmented shape provides a (continuous) measure of fracture severity.

2 Shape Estimation

The variations of vertebra shape over a training set of examples of unfractured spines are modeled using the linear point distribution models (PDM) as proposed by Cootes and Taylor [10]. PDMs model the shape probability distribution as a multivariate Gaussian in a subspace of reduced dimensionality. Shapes are defined by the coordinates of a set of landmark points which correspond between different shape instances. A collection of training shapes are aligned using for instance Procrustes analysis [11] and a principal component analysis (PCA) is applied to the aligned shape vectors. To this end, the mean shape $\overline{\mathbf{x}}$, the covariance matrix Σ, and the eigensystem of Σ are computed. The eigenvectors ϕ_i of Σ provide the so-called *modes of shape variation* which describe a joint displacement of all landmarks. The eigenvectors corresponding to the largest eigenvalues λ_i account for the largest variation; a small number of modes usually

captures most of the variation. Each shape $\mathbf{x}$ in the set can then be approximated by a linear combination of the mean shape and these modes of variation:

$$\mathbf{x} = \overline{\mathbf{x}} + \mathbf{\Phi_t b} + \mathbf{r}$$

where $\mathbf{\Phi_t}$ consists of the eigenvectors corresponding to the t largest eigenvalues, $\mathbf{\Phi_t} = (\phi_1|\phi_2|\ldots|\phi_t)$, $\mathbf{b}$ is a vector of model parameters that weigh the contribution of each of the modes, and $\mathbf{r}$ is a vector of residual shape variation outside of the model subspace.

2.1 Modeling Relations Between Shapes

The distribution $P(S_1|S_2)$, the probability distribution of a shape S_1 given a known other shape S_2, can be modeled as the Gaussian conditional density

$$P(S_1|S_2) = \mathcal{N}(\mu, K) \tag{1}$$

with

$$\mu = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(S_2 - \mu_2)$$

$$K = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

where μ_1 and μ_2 are the mean shapes of the training sets for S_1 and S_2, and covariances Σ_{ij} are obtained from the combined covariance matrix

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

as

$$\Sigma_{ij} = \frac{1}{n-1}\sum_n (S_{in} - \mu_i)(S_{jn} - \mu_j)^T.$$

$\Sigma_{12}\Sigma_{22}^{-1}$ is the matrix of regression coefficients of $S_1 - \mu_1$ on $S_2 - \mu_2$. Usually, Σ_{22} is not invertible owing to multi-collinearity in the landmark positions and unreliable due to chance covariance in a limited training set. Some regularization is therefore required. One option is to replace Σ_{22} by $\Sigma_{22} + \gamma I$, where γ is a positive and typically small constant. This approach is known as ridge regression [12]. As γ tends to infinity, the influence of the shape S_2 decreases, and the remaining model is the original model for S_1, describing the shape variation independent of S_2.

Applied to pairwise vertebra shape prediction from neighboring vertebrae, S_2 is the predictor vertebra shape, S_1 is the shape to predict, μ is the maximum likelihood estimate of S_1 given S_2, and K is the variance in the estimate.

One can choose to model both pose and shape with respect to the given vertebra, which means that the training shape pairs should be aligned together, on the basis of the transformations that optimally align the predictor vertebra in all training shapes. Alternatively, one could model the shape variation alone and leave out any possible correlations between shapes and relative position, scale, and rotation. In that case, the vertebrae should be aligned independently.

2.2 Estimator Uncertainty

The pairwise shape prediction results in several shape estimates for each vertebra. Not all of these estimates will be equally accurate. For instance, one would expect that the vertebral shape correlation between two direct neighbors is stronger than between two vertebrae that are further apart. In a fractured spine, the fractured vertebra(e) will likely produce inaccurate estimates of normal vertebra shape, even for its direct neighbors. We will therefore define the final shape estimate as a weighted combination of the individual predictions, where the weights express the degree of belief in each estimate.

We assume that the observed vertebral shapes are produced by the underlying shape model of normal shapes, resulting in a multi-variate Gaussian with variances λ_i in t directions, plus additional uncorrelated Gaussian noise with a variance σ_r^2 in all directions which accounts for any residual shape differences. The probability density for a shape S is then given by the product of the Gaussian densities of the shape model and the residual model:

$$p(S|\theta) = c_s c_r \exp[-\frac{1}{2}(\mathrm{M}_s + \mathrm{M}_r)] \tag{2}$$

$$c_s = \frac{1}{\sqrt{(2\pi)^t \prod_{i=1}^{t} \lambda_i}} \;,\; \mathrm{M}_s = \sum_{i=1}^{t} \frac{b_i^2}{\lambda_i}$$

$$c_r = \frac{1}{\sqrt{(2\pi)^n \sigma_r^{2n}}} \;,\; \mathrm{M}_r = \frac{|\mathbf{r}|^2}{\sigma_r^2}$$

where b_i are the model parameters from the PDM and $\mathbf{r}$ is a vector of residuals.

2.3 Combining Shape Estimates

The probability density for each conditional shape estimate can be expressed as $P(S_1|S_2)P(S_2)$, where $P(S_2)$ is the probability that the predictor shape S_2 is a valid normal (unfractured) shape, and the variance in $P(S_1|S_2)$ expresses the uncertainty in the prediction of S_1 from the model conditional on S_2. The weight for the ith prediction of S_1 is then given by

$$w_i = \frac{P(S_1|S_i)P(S_i)}{\sum_i P(S_1|S_i)P(S_i)}, \tag{3}$$

and the individual estimates are combined as a weighted sum.

$P(S_i)$ can be determined by substituting the predictor shape S_i for S in Equation 2 and the mean and covariance of the training set for the predictor shapes for θ, whereas for $P(S_1|S_i)$ the predicted shape for S_1 and the mean and covariance of the conditional model must be substituted. Note, that the regression models will typically produce shapes that are inside the model subspace — apart from some minor differences owing to the cut off at t eigenmodes — and thus in this case the residuals r_i and shape parameters b_i are negligible and the estimate reduces to a constant that is proportional to the inverse of the total variance in the conditional model. This constant is independent of both the predictor and the predicted shape and expresses the amount of correlation between the two shape models. In the pairwise vertebra predictions, the fact that direct neighbors contain the most useful information for predicting a shape is encoded in this term.

3 Vertebral Fracture Quantification

To summarize, we propose the following procedure for fracture quantification from segmented spine images:

- construct models of vertebral shape variation for all pairs of vertebrae in an image, according to Section 2
- for each pair of vertebrae in a new image:
  - align the predictor shape with the model
  - perform shape regression using Equation 1
  - determine reliability weights according to Equation 3
- combine all predictions for each vertebra as a weighted sum
- if pose is not included in the model, align prediction to true shape
- measure fracture severity

Various measures can be used to express the difference between the true shape and the predicted shape as a measure of vertebral deformity. In this work we ignore the points where the true shape is outside the prediction, which is usually caused by osteophytes. We then use the root mean squared (RMS) distance from all points where the prediction is outside the true shape to their closest points on the true shape.

4 Experiments

Our database currently consists of 212 lateral, lumbar spine radiographs taken from different programs of osteoporosis and atherosclerosis screening in post-menopausal women. The dataset is diverse, ranging from normal spines to spines with several severe fractures. The original radiographs have been scanned at a resolution of either 300 or 570 dpi and the lumbar vertebrae L1 — L4 were annotated and graded by experienced radiologists.

The outlines of the vertebrae were drawn manually and the corner points and midpoints of the vertebral endplates were indicated. Fractures were identified and graded according to the Genant at al. method of semi-quantitative visual assessment [1] in severity mild, moderate, or severe and type wedge, biconcave, or crush fracture. A total of 78 fractures was identified in 64 spines; 148 of the spines were unfractured.

A set of leave-one-out experiments is performed in which the models are trained on all available healthy vertebra shapes except those of the patient under study. To avoid including mild fractures in the model, shapes of which the difference between the maximum and minimum heights exceeds 15% of the maximum height are excluded from training.

4.1 Parameter Settings

A total of 52 landmarks is placed along the upper, anterior, and lower boundary of each vertebra, interpolated equidistantly between the four vertebral corners. The corners are defined as the points on the contour closest to the corner points that were used for standard six-point morphometry. Shapes are aligned individually using translation and rotation without scaling, since vertebra size is expected to correlate with important

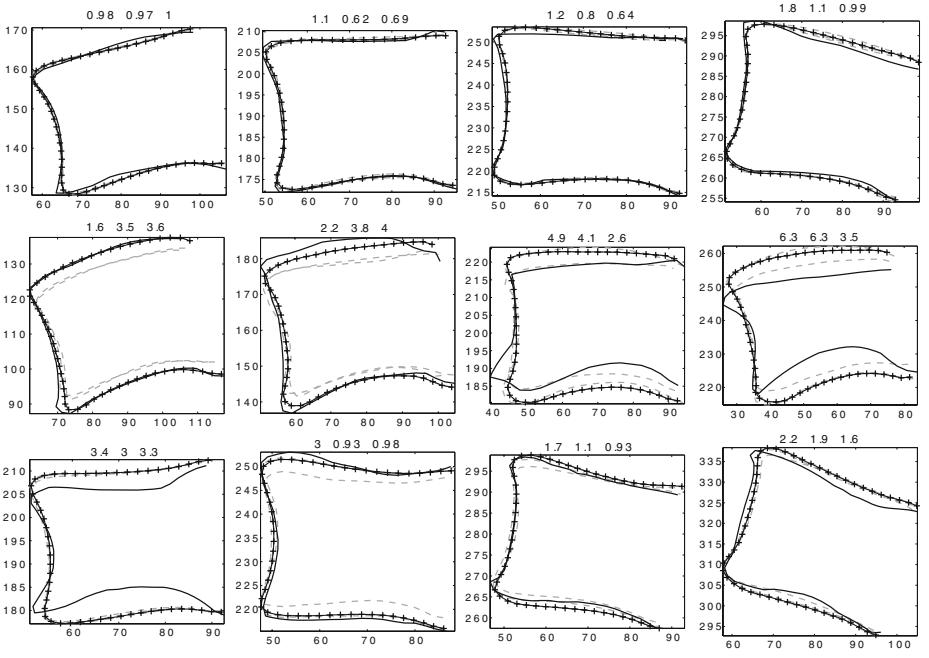


Fig. 1. Each row shows the vertebrae L1 — L4 of the same image; the true shape (black line), three pairwise predictions (gray dashed lines), and the combined prediction (black line with pluses). The numbers above the plots give the RMS distance to the true shape for each of the three individual predictions. The top row depicts a normal spine; row 2 contains 2 fractures (vertebra 3 and 4); row 3 was graded as all normals but vertebra 1 had a large shape difference in our method.

shape information on fracture type and severity. The number of modes is selected so that 95% of the variance is kept in the model. The residual variance and the regularization parameters γ were optimized separately for each pairwise regression model using leave-one-out validation on the training set of normal vertebrae.

4.2 Results

Some examples of predictions obtained are given in Figure 1. Most of the normal shapes are predicted accurately. The second row in Figure 1 contains two fractures, which results in two of the three predictions being consistently smaller than the true shape. However, the model is able to detect that those shapes are less likely and the weighted estimate for the normal vertebrae is still close to the correct shape.

Figure 2 summarizes the prediction results of all 212 images, separated into fracture type and grade. The RMS shape-to-contour distance between the predicted shape and the true contour was 1.0 mm on average for unfractured vertebrae and 3.67 mm for fractures. The area under the ROC curve is 0.93; at a sensitivity of 95% the specificity is 84%. The areas under the ROC curve for mild, moderate, and severe fractures separately are 0.92, 0.96. and 0.99 respectively.

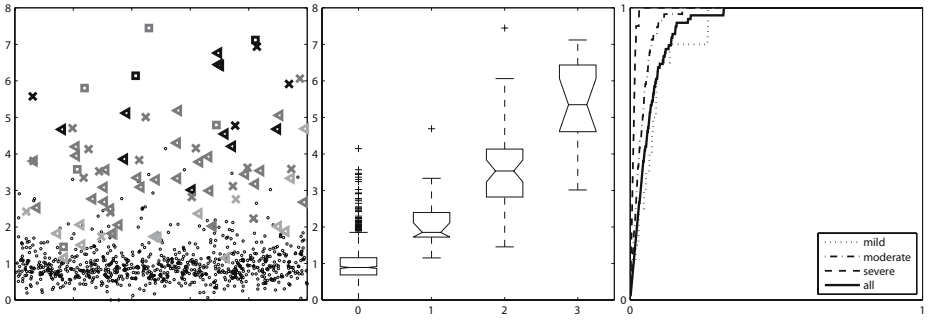


Fig. 2. (a) Measured RMS distances in mm between the predicted shapes and true shapes. To allow discrimination of similar measures the distances are set out against a random value on the horizontal axis. The symbols denote the type of vertebral deformity: Normal (circles); wedge fracture (triangles); biconcave fracture (crosses); and crush fracture (squares). For the fractures, the grey value denotes the fracture severity: mild (light grey); moderate (darker grey); and severe (black) **(b)** Matlab box and whisker plot of the average distances between predicted and true shapes separated into fracture grade, ranging from normal (0) to severely fractured (3). The box has lines at the lower quartile, median, and upper quartile values. The whiskers extending from each end of the box show the extent of the rest of the data. Data with values outside 1.5 times the interquartile range from the median are considered to be outliers and are denoted by pluses.**(c)** ROC curves of vertebral fracture detection.

Overall, we are able to distinguish fractures from normals and there is a fair correlation (0.81) between the shape distance and the fracture severity as indicated by the radiologists. However, Figure 1 reveals that the class of normal vertebrae contains a relatively large number of outliers that have a larger difference between the prediction and the true shape. One of these outliers is shown in the third row of Figure 1. The leftmost vertebra may be a very mild 'fracture' that falls outside the capture range of the standard semi-quantitative morphometry and that the shape prediction, more sensitive to subtle changes, can already identify. This should be further validated using for instance longitudinal studies. Visual inspection of the dataset revealed that many of the unfractured vertebrae that had a large shape distance exhibit this type of deformation.

5 Discussion and Conclusion

We propose a shape model based approach to vertebral fracture quantification in which an observed vertebral shape is compared to its reconstructed normal shape as can be predicted from its neighbors. Compared to the current standard of semi-quantitative morphometry which is based on three height measurements per vertebra, this method provides a richer description of deformation and may be able to detect more subtle shape changes while maintaining specificity. This could lead to earlier diagnosis and reduce the number of participants and/or the follow-up time required in clinical trials assessing the efficacy of drug candidates.

Currently, the shape differences are summarized into one average distance. This enables detection of deformities and specification of the degree of abnormality, but not

the type. If a sufficiently large training set of different types of deformity is available, it may be useful to take the shape difference vectors — or a low-dimensional representation of them — as input in a supervised classification scheme. Such explicit models of deformities could also be incorporated in the step of combining different shape estimates, such that the estimate of uncertainty is not based only on the distance to known normal shapes but also on the distance to known deformities.

In the current setup, vertebral shapes were reconstructed with an accuracy of on average 1.0 mm for normal vertebrae and 3.7 mm for fractures, leading to fracture detection with a sensitivity of 95% at a specificity of 84%.

References

1. Genant, H.K., Wu, C.Y., van Kuijk, C., Nevitt, M.C.: Vertebral fracture assessment using a semiquantitative technique. J Bone Miner Res **8**(9) (1993) 1137–1148
2. Genant, H.K., Jergas, M., Palermo, L., Nevitt, M., Valentin, R.S., Black, D., Cummings, S.R.: Comparison of semiquantitative visual and quantitative morphometric assessment of prevalent and incident vertebral fractures in osteoporosis The Study of Osteoporotic Fractures Research Group. J Bone Miner Res **11**(7) (1996) 984–996
3. McCloskey, E.V., Spector, T.D., Eyres, K.S., Fern, E.D., O'Rourke, N., Vasikaran, S., Kanis, J.A.: The assessment of vertebral deformity: a method for use in population studies and clinical trials. Osteoporos Int **3**(3) (1993) 138–147
4. Ferrar, L., Jiang, G., Adams, J., Eastell, R.: Identification of vertebral fractures: an update. Osteoporos Int **16**(7) (2005) 717–728
5. Smyth, P., Taylor, C., Adams, J.: Vertebral shape: Automatic measurement with active shape models. Radiology **211**(2) (1999) 571–578
6. Zamora, G., Sari-Sarrafa, H., Long, R.: Hierarchical segmentation of vertebrae from X-ray images. In Sonka, M., Fitzpatrick, M., eds.: Med Imaging: Image Process. Volume 5032 of Proc of SPIE., SPIE Press (2003) 631–642
7. de Bruijne, M., Nielsen, M.: Image segmentation by shape particle filtering. In Kittler, J., Petrou, M., Nixon, M., eds.: ICPR, IEEE Computer Society Press (2004) III:722–725
8. Roberts, M.G., Cootes, T.F., Adams, J.E.: Vertebral shape: Automatic measurement with dynamically sequenced active appearance models. In Duncan, J., Gerig, G., eds.: MICCAI. Volume 3750 of LNCS., Springer (2005) 733–740
9. Lund, M., de Bruijne, M., Tankó, L., Nielsen, M.: Shape regression for vertebra fracture quantification. In Fitzpatrick, M., Reinhardt, J., eds.: Med Imaging: Image Process. Volume 5747 of Proc of SPIE., SPIE Press (2005) 723–731
10. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active shape models – their training and application. Comput Vis Image Underst **61**(1) (1995) 38–59
11. Goodall, C.: Procrustes methods in the statistical analysis of shape. Journal of the Royal Statistical Society B **53**(2) (1991) 285–339
12. Hoerl, A., Kennard, R.: Ridge regression: Biased estimation for nonorthogonal problems. Technometrics **12**(1) (1970) 55–67

Anatomically Constrained Deformation for Design of Cranial Implant: Methodology and Validation

Ting Wu[1], Martin Engelhardt[2], Lorenz Fieten[1],
Aleksandra Popovic[1], and Klaus Radermacher[1]

[1] Helmholtz Institute for Biomedical Engineering, RWTH Aachen University, Aachen, Germany
{twu, fieten, popovic, radermacher}@hia.rwth-aachen.de
[2] Clinic for Neurosurgery, Ruhr-University, Bochum, Germany
martin.engelhardt@ruhr-uni-bochum.de

Abstract. We present a new approach for cranial implant design which uses anatomical constrained deformation based on reference models. The methodological framework contains three steps: patient-specific generation of the reference model containing the anatomical constraints about the skull shape; determination of the spatial correspondence between the patient skull and the reference model by 3D matching; adaptive deformation of the fragment on the reference model corresponding to the defect area on the patient skull for implant design. The proposed method was validated by simulating the reconstruction of artificially generated defects on healthy skulls. The validation results show that this approach can generate implant geometry very fast and with satisfactory quality. This approach also outperforms the surface interpolation method in reconstructing cranial defects.

1 Introduction

Calvarial tumors, infectious bone flap, and neurosurgical external decompression are the main causes for large cranial defects. The main indications for cranial reconstruction of these patients are cosmetic reasons and protection of intracranial structures from mechanical impact [1][2]. With the advance of the three-dimensional (3-D) Computed Tomography (CT) it is possible to reconstruct the 3D patient anatomy and design an implant with help of computers. The currently well-established approaches for implant design often involve a complicated CAD/CAM process chain [3][4][5][6]. In the CAD/CAM based approaches a surgeon needs support from technical staffs or tools out of clinics, which sometimes leads to high cost and low efficiency. Despite the technical improvement in CT and CAD/CAM, the reconstruction of a large-format skull defect sometimes remains difficult because manual construction and modelling of a free-form surface is necessary for each individual patient. Some alternative methods have been proposed to design the implant based on CT data without CAD process by using methods like mirroring [7], surface interpolation or deformation [8][9]. These methods use either

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 9 – 16, 2006.

anatomical or mathematical features of the skull surface to reconstruct the defect area. The mirroring method uses the anatomical features on the contra-lateral part but is suitable for the defect on one side of the skull (unilateral defect) only. The interpolation and deformation methods, such as thin-plate spline (TPS), use mathematical constraints (e.g. C^1 continuity) to design the implant from the remaining part of a skull, however, without considering anatomical constraints.

In this paper, we present a novel approach for cranial implant design which uses anatomically constrained adaptive deformation based on a reference model. The patient-specific reference model is used as anatomical constraints for implant design. The approach was validated by simulating the reconstruction of 44 artificial defects on nine healthy skulls with known ground truth. Furthermore, we also compare the performance of the proposed method with the surface interpolation method.

2 Methods

The methodological framework of anatomically constrained deformation for implant design can be divided into three steps (Fig. 1.): Firstly, a reference model is generated according to the patient skull; then 3D-matching is performed to determine the spatial correspondence between the patient skull and the reference model; finally, adaptive deformation is used to design the implant geometry according to the defect area on the patient skull and its corresponding fragment on the reference model.

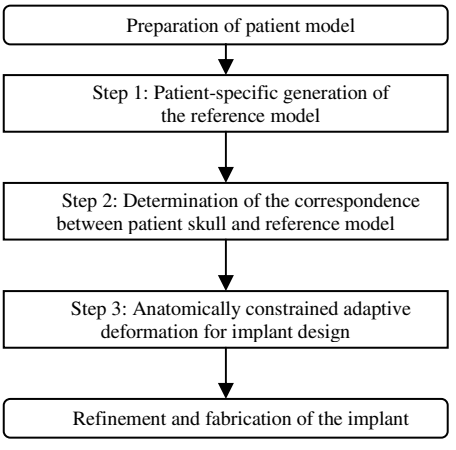


Fig. 1. The methodological framework for anatomically constrained implant design based on reference models in the workflow of cranial implant design and fabrication

2.1 Patient-Specific Generation of the Reference Model

The reference model can be generated in different ways according to the patient skull defect. For unilateral defects, mirroring the normal side could be used to

generate the reference model. For cranial defects spanning both sides of the skull (bilateral defects), or more generally (i.e. also applied to unilateral cases), the reference model can be obtained by searching a similar skull in a 3D reference database of healthy skulls. The reference database contains currently 25 healthy skulls (13 Male, 12 Female) with mean age 36 and median age 39 (range 20-62 years). To enable the efficient 3D retrieval of the similar skull model, the original datasets are indexed by descriptors about the demographic information (e.g. age, sex, etc.) and geometrical characters based on anthropometric measurement of craniofacial landmarks [10]. The craniofacial landmarks are also used to define standardized cutting planes or view directions. The contour of a 2D images on such planes or in such view directions is then extracted and represented by morphological descriptors based on normalized Fourier descriptors, which enable a similarity comparison with scaling and rotation invariance.

2.2 Determination of the Spatial Correspondence by 3D Matching

The paired point matching method [11] was used to match craniofacial landmarks on the patient skull and the corresponding landmarks on the reference model. It is also possible to do the matching based on the user-defined points on the skull surface through user interaction. Through geometric matching, a spatial correspondence between the points of the patient skull and the points of the reference model is established. The aim of this step is to obtain an estimation of transformational relationship between the patient skull and the reference model. Since the shapes of the patient skull and the reference model might be totally different, surface-based matching methods like ICP [12] are not used here. Moreover, the paired point matching method is computationally more efficient for our purpose.

2.3 Implant Design Based on Anatomically Constrained Adaptive Deformation

The main idea of the proposed method is to use the reference model as anatomical constraint to control the deformation adaptively at each point. An internal coordinate system based on the craniofacial landmarks is defined so that a point on the surface of the patient skull and the reference model can be described in spherical form $r_{pat}(\varphi,\theta)$ and $r_{ref}(\varphi,\theta)$ (i.e. radial distance as function of the directional angles: φ and θ). The origin of the coordinate system is defined as the centroid of the landmark pair Porion (po, the most superior point on the upper margin of the external auditory meatus when the head is in the Frankfort horizontal plane). The directions of the three axes in Cartesian coordinate system are defined as follows: the average of the directional vectors of the 2 connecting lines between landmarks Porion (po) and Frontozygomaticus (fz, the most lateral points on the frontozygomatic suture) on the left and on the right side, the directional vector along the connecting line from Porion (po) on the right side to Porion (po) on the left side, and their cross product [13].

The implant geometry is generated by adaptive deformation based on radial scaling. The direction-dependent radial scaling factor $s(\varphi,\theta)$ is defined as:

$$s(\varphi,\theta) = r_{pat}(\varphi,\theta) / r_{ref}(\varphi',\theta') \tag{1}$$

where the φ and θ are the directional angle in the coordinate system of the patient skull and the φ' and θ' are the corresponding directional angle in the coordinate system of the reference model calculated by using the transformational relation determined in the step 2. By user-interactive delineation of the defect margin, a set of points on the patient skull $\{ r_{pat}(\varphi_i,\theta_i),\ i=1, \ldots, N\}$ are defined. Since the radial distance of their corresponding points on the reference model can be determined, the scaling factor $s(\varphi_i,\theta_i)$ at each point $r_{pat}(\varphi_i,\theta_i)$ is known. Taking all these points as control points, the scaling factor $s(\varphi,\theta)$ on the defect area of the patient skull (i.e. where $r_{pat}(\varphi,\theta)$ is unknown) can be estimated by interpolation according to directional angles. Since the radial distance on the fragment of the reference model corresponding to the defect area of the patient skull is known, the radial distance at the defect area can be estimated by scaling the radial distance on the reference model with the interpolated scaling factor. In this way, the coordinates of the surface points of the implant to reconstruct the defect can be calculated. So the implant surface is the result of deforming the corresponding fragment on the reference model by radial adaptive scaling. The thickness of the implant can be set by users. In contrast to the direct surface interpolation method, we first interpolate the scaling factor and then multiply the scaling factor with the radial distance on the reference model so that both the remaining part of the patient skull and the reference model (which contains anatomical constraints) contribute to the final implant geometry (Fig. 2.).

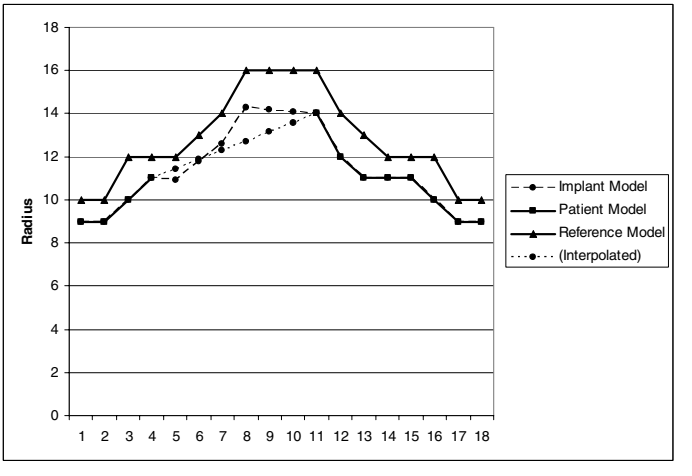


Fig. 2. Illustration of the anatomically constrained deformation. The reference model gives the anatomical constrain for implant geometry while direct interpolation fails to recover certain anatomical detail on the defect area.

Further refinement such as morphological operation is used to smooth the implant and make it fit well with the defect on the patient skull.

3 Results

To validate the proposed methods, nine datasets of healthy skull were selected and totally 44 defects with different size and location were generated from these datasets. Since the ground truth is known, we can evaluate the quality of the reconstruction. We also compared the proposed method with the surface interpolation method.

3.1 Computational Performance

The time needed for the entire reconstruction process is less than 25 minutes on a Pentium 4 computer (3.2 GHz, 1GB RAM). Compared with the CAD/CAM based approaches which might need a couple of days, it is now possible for the surgeon to perform the planning task of implant design on his own.

3.2 Validation of the Reconstruction Quality with Ground Truth

The reconstruction quality was evaluated by comparing the healthy skull (ground truth) and the test datasets (i.e. the original healthy skull with an artificial defect) reconstructed with the implant designed with the proposed method. Taking 9 healthy skulls (5 Male, 4 Female, Age range 26-47) as ground truth, 44 test datasets are generated by artificially removing a fragment from one of the healthy skulls. These 9 healthy skulls were not contained in the reference database. The defects have different sizes and different locations. The assessment criterion of the reconstruction quality is the reconstruction error defined as the difference between radial distance of surface voxels on the original skull (ground truth) and radial distance of surface voxels on the reconstructed part of the test datasets (i.e. implant). The reconstruction error is measured in voxels so that this criterion is independent of the resolution of CT. The average percentage of surface voxels with reconstruction error within certain ranges (e.g. error < 1 voxel) were calculated for statistical evaluation.

Table 1. The reconstruction quality of the implant measured by the difference in the radial distance: the distribution of the reconstruction error within different range

	0-1 voxel	2-4 voxels	>4 voxels
Total (44)	81.6% (68%-96%)	12.8% (4%-19%)	5.6 % (0%-13%)
Group A (21)	82.8%	12.0%	5.2%
Group B (23)	80.5%	13.5%	6.0%
Group 1 (16)	88.0%	9.0%	3.0%
Group 2 (28)	77.9%	15.1%	7.1%

An average of 81.6% (68% - 96%) of the surface points can be reconstructed within an error of 1 voxel. The average percentage of surface points reconstructed

with large error (> 4 voxels) is less than 6% (Table 1.). According to the defect size, the test datasets are subdivided into 2 groups: Group A covering less than 100 cm^2 (n=21), Group B covering more than 100 cm^2 (n=23). There is no significant difference between these two groups considering the reconstruction quality. The test datasets can also be subdivided into 2 groups according to coverage and location: Group 1 with unilateral defect (n=16) and Group 2 with bilateral defect (n=28). There is a significant difference between these two groups because the mirrored model provides more patient-specific and precise anatomical constraints for implant design. This can be further confirmed when implants are designed using reference skull models instead of mirrored models for unilateral defects (Table 2.).

Table 2. Comparison of reconstruction quality of the implant for unilateral defects (Total 16) using mirrored models (Method 1) and similar skulls as reference models (Method 2)

	0-1 voxel	2-4 voxels	>4 voxels
Method 1	88.0%	9.0%	3.0%
Method 2	79.5%	14.8%	5.7%

It is also shown that the reconstruction quality is strongly dependent on the availability of similar skulls in the reference database when reconstructing bilateral defects (Table 3.). However, if the reference database contains skulls with large varieties, good results could be expected in general.

Table 3. Comparison of reconstruction quality of the implant for skull defects (Total 20) using the most similar skull (Method 1) and using a skull by randomized selection in the reference skull database (Method 2) as the reference model. Method 2 has a large deviation in reconstruction error compared with Method 1.

	0-1 voxel	2-4 voxels	>4 voxels
Method 1	79.6% (68%-93%)	14.1% (6%-19%)	6.3% (1%-13%)
Method 2	62.3% (38%-87%)	27.2% (10%-46%)	10.5% (3%-29%)

3.3 Comparison to Surface Interpolation Methods

The proposed method has been compared with the surface interpolation method using thin-plate splines [8]. Twelve test datasets were reconstructed by these two methods. It is shown that our method outperforms the surface interpolation method. Although the surface interpolation method can recover a smooth surface on the defect area, it has no anatomical constraints which can contribute to a better reconstruction quality for individual patients (Fig.3. and Table 4.). For the surface interpolation method, the surface voxels near to the defect margin have usually less reconstruction error whereas the surface voxels at the middle part of the defect area have often the large reconstruction error. This shows the limitation of pure mathematical constraints for implant design.

Table 4. Comparison of reconstruction quality of 12 test datasets between the proposed method (Method 1) and the surface interpolation method (Method 2)

	0-1 voxel	2-4 voxels	>4 voxels
Method 1	81.0%	13.1%	5.9%
Method 2	69.5%	18.8%	11.7%

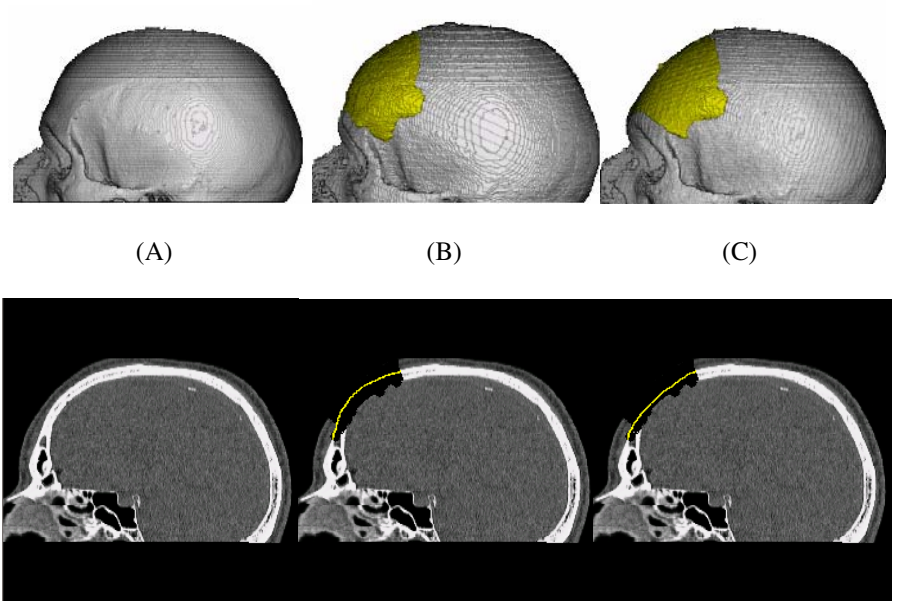


Fig. 3. Comparison of the reconstruction using anatomically constrained deformation and surface interpolation with thin-plate splines. (A) Original dataset (ground truth) (B) Reconstruction with anatomically constrained deformation (the yellow part is implant) (C) Reconstruction with thin-plate splines (D) Midsagittal slice of the original dataset (E) Midsagittal slice of reconstruction with anatomically constrained deformation (F) Midsagittal slice of reconstruction with thin-plate splines.

4 Conclusions

In this paper, a new approach to design cranial implant using anatomically constrained deformation is presented and validated by reconstruction simulated skull defects with known ground truth. The results show that the proposed method is fast and can design implants with quite good quality. Therefore, the surgeon can perform the task of designing the implant geometry without the help of technicians and use the original CT data for medical verification in one planning session.

Besides the clinical studies, the ongoing work includes introducing statistical shape models of skull in this methodological framework and investigation of the influence of the user-interactive localization of craniofacial landmarks and the defect margin to the reproducibility of the implant design. Furthermore, the database structure and its search strategies will be extended and optimized for an internet-based architecture so that the central reference database of the skull geometry can be shared by different users.

References

1. Sanan A, Haines SJ.: Repairing holes in the head: a history of cranioplasty. Neurosurgery **40** (1997) 588–603
2. Dujovny M, Aviles A, Agner C, Fernandez P, Charbel FT.: Cranioplasty: cosmetic or therapeutic? Surg Neurol **47** (1997) 238-241
3. Eufinger H, Wehmöller M, Machtens E, Heuser L, Harders A, Kruse D. Reconstruction of craniofacial bone defects with individual alloplastic implants based on CAD/CAM-manipulated CT-data. J Cranio Maxillofac Surg **23** (1995) 175-181
4. Heissler E, Fischer FS, Bolouri S, Lehmann T, Mathar W, Gebhardt A, et al. Custom-made cast titanium implants produced with CAD/CAM for the reconstruction of cranium defects. Int J Oral Maxillofac Surg **27** (1998) 334–338
5. Lee MY, Chang CC, Lin CC, Lo LJ, Chen YR. Custom implant design for patients with cranial defects. IEEE Eng Med Biol Mag **21**(2) (2002) 38-44
6. Müller A, Krishnan KG, Uhl E, Mast G. The application of rapid prototyping techniques in cranial reconstruction and preoperative planning in neurosurgery. J Craniofac Surg **14** (2003) 899-914
7. Sauret V, Linney AD, Richards R. Computer assisted surgery: the use of digital images in enabling computerized design and manufacture of titanium implants. Imaging **14** (2002) 464-471
8. Carr JC, Fright WR, Beatson RK. Surface interpolation with radial basis functions for medical imaging. IEEE Transactions in Medical Imaging **16**(1) (1997) 96-107
9. Min KJ., Dean D.: Highly accurate CAD tools for cranial implants. In Ellis, R.E. and Peters, T.M., eds.: Proceedings of Sixth International Congerence on Medical Image Computing and Computer Assisted Intervention, MICCAI 2003. Volume 2878 of Lecture Notes on Computer Science., Springer-Verlag (2003) 99-107
10. Kolar JC, Salter EM. Craniofacial anthropometry practical measurements of the head and face for clinical, surgical and research use. Charles Thomas publisher ltd., USA (1996).
11. Horn BKP. Closed-form solution of absolute orientation using unit quaternions. J. Opt. Soc. Am. A **4**(4) (1987) 629-642
12. Besl PJ, McKay ND. A method for registration of 3-d shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence **14**(2) (1992) 239-256

Open-Curve Shape Correspondence Without Endpoint Correspondence

Theodor Richardson and Song Wang

Department of Computer Science and Engineering,
University of South Carolina, Columbia, SC 29208, USA
richa268@cse.sc.edu, songwang@cse.sc.edu

Abstract. Shape correspondence is the foundation for accurate statistical shape analysis; this is usually accomplished by identifying a set of sparsely sampled and well-corresponded landmark points across a population of shape instances. However, most available shape correspondence methods can only effectively deal with complete-shape correspondence, where a one-to-one mapping is assumed between any two shape instances. In this paper, we present a novel algorithm to correspond 2D open-curve partial-shape instances where one shape instance may only be mapped to part of the other, i.e., the endpoints of these open-curve shape instances are not presumably corresponded. In this algorithm, some initially identified landmarks, including the ones at or near the endpoints of the shape instances, are refined by allowing them to slide freely along the shape contour to minimize the shape-correspondence error. To avoid being trapped into local optima, we develop a simple method to construct a better initialization of the landmarks and introduce some additional constraints to the landmark sliding. We evaluate the proposed algorithm on 32 femur shape instances in comparison to some current methods.

1 Introduction

It is well known that the performance of statistical shape analysis [1, 3, 10, 7, 8] is highly dependent upon the performance of shape correspondence [4, 2, 11, 9], which identifies a set of sparsely sampled and well-corresponded landmark points across a population of shape instances. However, accurate shape correspondence is a very challenging problem given the strong nonlinearity of shape geometry and the complex nonrigid deformation occurring among the considered shape instances.

Many new models and methods have been developed recently to achieve more accurate landmark-based shape correspondence. Each method has its own merits and difficulties. Davies et al [4] develop a Minimum Description Length (MDL) algorithm where the correspondence error is measured by the required bit-length to transmit these shape instances as well as the template shape. Both random-searching [4] and gradient-descent algorithms [6] have been developed to minimize this nonlinear MDL-based measure. Xie and Heng [14] develop a medial-axis-based shape-correspondence method, where the medial axes of all

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 17–24, 2006.

the shape instances are assumed to be of the same topology and then the shape-correspondence problem is decomposed into a set of simpler problems of corresponding some short curve segments with very few high-curvature points. With some roughly-corresponded landmarks, Bookstein [2] develops an algorithm to move these landmarks along the tangent directions of the shape contour to achieve a minimum landmark-correspondence error that is defined by the thin-plate bending energy. However, the resultant landmarks may not be located on the underlying shape contour. Wang, Kubota, and Richardson [13, 9] address this problem by adding a step of projecting the landmarks back to the shape contour.

However, all these methods can only address the complete-shape correspondence where a one-to-one mapping is assumed between each pair of the shape instances. In 2D cases, each shape instance is in the form of a continuous curve. For convenience, we sometimes refer to this ground-truth continuous form of a shape instance as a *shape contour*. These shape contours may be open or closed (the starting and ending point of the contour is in the same physical location). When the shape contour is open, the endpoints are always corresponded landmarks across all the shape instances. In many real applications, however, the assumption of the complete-correspondence does not usually hold for open-curve shape contours. An shown in the femur contours in Fig. 1, we can see that it is difficult to guarantee that the initially extracted open femur boundaries have exact endpoint correspondence and one boundary may only be mapped to a portion of the other. Such examples are common in practice since we may only be interested in portions of a structure and it is difficult, even manually, to accurately extract open-curve shape contours with consistently corresponded endpoints. Furthermore, many medical images suffer from cropping and occlusion within a whole structure. In this paper, we develop an algorithm to address such an open-curve partial shape correspondence problem by extending and adapting the landmark-sliding algorithms [2, 13, 9]. This algorithm is tested on 32 femur shape instances and the correspondence performance is quantitatively compared with the results of a popular implementation of the MDL algorithm [12] and the most recent landmark-sliding algorithm [9].

2 Problem Formulation and Landmark-Sliding

In this paper, we consider 2D partial-shape correspondence, where each shape instance is in the form of a continuous open-shape contour, as shown in Fig. 1. Denote the given set of open-curve shape instances to be $S = \{S_1, S_2, \ldots, S_n\}$. Each shape instance S_i is in the form of an arc-length parameterized curve $\mathbf{s}_i(t_i) = (x_i(t_i), y_i(t_i))$, $0 \leq t_i \leq L_i$, where L_i is the perimeter length of S_i and t_i is the traversed arc-length along the curve from $\mathbf{s}_i(0)$ to $\mathbf{s}_i(t_i)$. Note that $\mathbf{s}_i(0)$ and $\mathbf{s}_i(L_i)$ are the two endpoints of the shape contour S_i. In practice, we obtain these shape instances by applying the Catmull-Rom spline interpolation to approximate some structural boundaries that are manually extracted from the medical images.

To address the partial shape correspondence, we need to determine a continuous subcontour P_i of each shape contour S_i such that a complete correspondence exists across P_i, $i = 1, \ldots, n$. As a subcontour of S_i, we can denote the two endpoints of P_i as $P_{i1} = \mathbf{s}_i(t_{i1})$ to $P_{iK} = \mathbf{s}_i(t_{iK})$, where $t_{i1} \geq 0$ and $t_{iK} \leq L_i$ are the traversed curve lengths from $\mathbf{s}_i(0)$ to these two endpoints of P_i. Without loss of generality, we let $t_{i1} < t_{iK}$. The problem formulation of the open-curve complete-shape correspondence is well known for nonrigid shape correspondence: identifying K nonfixed landmarks along P_i, $i = 1, \ldots, n$ such that a predefined correspondence error among them is minimum and the two endpoints of each shape contour are always selected as corresponded landmarks, i.e., P_{i1}, $i = 1, \ldots, n$ are priorly known to be corresponded and P_{iK}, $i = 1, \ldots, n$ are also priorly known to be corresponded. This way, we can define the open-curve partial shape correspondence as finding parameters $0 \leq t_{i1} \leq \ldots \leq t_{iK} \leq L_i$, $i = 1, \ldots, n$ such that the sampled landmarks $\mathbf{s}(t_{ik})$, $k = 1, \ldots, K$, $i = 1, \ldots, n$ minimize a selected correspondence error.

As in [2] and [13, 9], we define the landmark shape-correspondence error using the thin-plate bending energy [5], which has been widely used to describe nonrigid shape deformation. Since corresponding n shape instances is usually treated by corresponding each shape instance separately to a template shape instance, we can focus on the correspondence between two shape instances: the template V in which the K sequentially sampled landmarks $\mathbf{v}_1, \ldots, \mathbf{v}_K$ have been priorly given and the target U in which we want to find the K corresponded landmarks $\mathbf{u}_1, \ldots, \mathbf{u}_K$. The correspondence error is defined to be the thin-plate bending energy

$$\beta(V \to U) = \frac{1}{8\pi}(\mathbf{x}_u^T \mathbf{M}_v \mathbf{x}_u + \mathbf{y}_u^T \mathbf{M}_v \mathbf{y}_u),$$

where $\mathbf{x}_u$ and $\mathbf{y}_u$ are columnized vectors of x- and y-coordinates of K landmarks along the target U, and $\mathbf{M}_v$ is the thin-plate bending matrix calculated from K landmarks in the template V [5]. Specifically, the bending matrix $\mathbf{M}_v$ is the upper-left $K \times K$ submatrix of

$$\begin{pmatrix} \mathbf{K}_v & \mathbf{P}_v \\ \mathbf{P}_v^T & \mathbf{0} \end{pmatrix}^{-1},$$

where $\mathbf{K}_v$ is $K \times K$ kernel matrix with element $k_{ij} = \mathcal{K}(\mathbf{v}_i, \mathbf{v}_j) = \|\mathbf{v}_i - \mathbf{v}_j\|^2 \log \|\mathbf{v}_i - \mathbf{v}_j\|$ and $\mathbf{P}_v = (\mathbf{x}_v, \mathbf{y}_v, \mathbf{1})$.

Given the nonlinearity of the continuous shape contour and the inability to slide landmarks along the contour directly, we adopt the landmark sliding method of [13], which starts from some initialized K landmarks along the target shape U, moves the landmarks along the tangent directions to reduce the correspondence error, and then projects the updated landmarks back to the shape contour. The sliding and projection operations are iteratively performed until convergence. Specifically, letting α_k be the sliding distance of $\mathbf{u}_k$, $k = 1, \ldots, K$, the landmark-sliding algorithm aims to minimize the correspondence error

$$\phi(\alpha_1, \ldots, \alpha_K) = (\mathbf{x}_u + \mathbf{P}'_x \boldsymbol{\alpha})^T \mathbf{M}_v (\mathbf{x}_u + \mathbf{P}'_x \boldsymbol{\alpha}) + (\mathbf{y}_u + \mathbf{P}'_y \boldsymbol{\alpha})^T \mathbf{M}_v (\mathbf{y}_u + \mathbf{P}'_y \boldsymbol{\alpha}) \quad (1)$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)^T$ and $\mathbf{P}'_x$ and $\mathbf{P}'_y$ are diagonal matrices of the x- and y-components of the estimated tangent vectors at the current K landmarks in the target shape U. The projection is performed by simply traversing the current landmarks along the shape contour by the sliding distance α_k, $k = 1, \ldots, K$. To preserve the shape topology, each landmark is not allowed to slide across its two neighbors; in [13], it is shown that this can be described by some linear constraints [13, 9]. Therefore, the above cost function can be optimized efficiently by the quadratic-programming algorithm. The major problem of the landmark sliding algorithm is its vulnerability to being trapped into local minima. For complete-shape correspondence with known corresponded endpoints, good initialization of the K target landmarks can be easily constructed, which, however, is not the case in the partial-shape correspondence discussed in this paper.

3 Landmark-Sliding for Open-Curve Partial Shape Correspondence

There are several important issues in extending the landmark-sliding algorithm to the open-curve partial shape correspondence. First, the landmark initialization becomes a very important issue without endpoint correspondence. Second, by moving the first and last landmark points, we actually resize the target shape to be mapped to the template. Reducing the size of the target shape, although it is an affine transformation, can decrease the bending energy. Therefore, without further considerations, we can not directly use the bending energy to decide the desirable correspondence results. Third, we also cannot allow the landmarks to move beyond the original endpoints of the open shape contours.

To simplify the algorithm description, we assume that the template shape contour V corresponds to part of the target shape U, i.e., the endpoints of V can be included as landmarks and their correspondence in U is located along U. Then the third issue above is easy to handle: we need only to add a constraint on the sliding distance α_1 and α_K so that they are smaller than the respective remaining external curve length. These are still linear constraints and, therefore, the quadratic programming algorithm still applies. In the following, we extend the landmark-sliding algorithms to address the other two important issues.

3.1 Initial Landmark Construction

The landmark-sliding algorithm is vulnerable to being trapped into a local optima, especially when the considered shape instances contain many high-curvature points, motivating an accurate initialization method. In the open-curve partial shape correspondence, it can be assumed that the shape contours extracted from medical images contain a similar subject of interest that can be extracted within a certain variance expressed as a percentage. For completeness, this variance is not assumed to be *a priori* knowledge though it may be known by an expert. For the cases considered herein, it is reasonable to assume that the variance of shape information at the endpoints is not greater than 50 percent.

We therefore iteratively test the bending energy by clipping an equal percentage ω from both sides of target shape instance U and distributing the K landmarks as described above over the remainder of the curve as follows:

$$\mathbf{t}_u = \mathbf{t}_v \frac{(1-\omega)L_u}{L_v} + \frac{\omega L_u}{2} \tag{2}$$

where $\mathbf{t}_u$ is the set of arc-length parameters for the K landmarks of shape instance U and L_u is the perimeter length of shape instance U ($\mathbf{t}_v$ and L_v are similarly defined for template shape instance V). With each distribution, we can then test the initial bending energy to determine the optimal ω clip length that should be used to establish the final correspondence with V. Because the landmarks of each U are not fixed, the step length is flexible and, within a certain range, the sliding will converge to the same result; for our purposes, we varied the clip length from 0 to 50 percent with a step length of 0.5 percent. The optimal clip length for each shape instance is determined independently to account for greater variability. As an added constraint to prevent undue collapse of the shape contour, endpoints are not allowed to move inward more than $\frac{\omega L_u}{2}$ for each shape instance U; however, the endpoints are allowed to move outward to recapture length that was previously assumed to be removed.

3.2 An Adapted Correspondence Error

The remaining problem is to choose the desirable correspondence result out of the ones that are resulting from different clip lengths. It is well known that the thin-plate bending energy is invariant to the affine transformation, i.e., if the map between V and U can be described by an affine transformation, the resulting bending energy is always zero. However, if the map between V and U is not an affine transformation, then some additional affine transformations on the target landmarks U would change the bending energy that is calculated without consideration of this additional affine transformation. As an example, a scaling factor of 0.75 applied to the target landmarks U in Eq. (2) would decrease the bending energy $\beta(V \rightarrow U)$ to $(0.75)^2\beta(V \rightarrow U)$. A scaling in either x- or y-coordinates alone will decrease the bending energy in a similar way. In a degenerate case where all target landmarks are moved to the same point, the size of the target shape contour that finally corresponds to the template is zero and, in this case, the bending energy, or the shape-correspondence error, is also zero. Therefore, the direct comparison of the bending energy may not help us decide the desirable correspondence result out of the ones from different clip lengths.

To address this problem, we introduce a normalized bending energy to help decide the desirable correspondence result. Specifically, for the correspondence result from each clip length, we normalize the resulting U by removing the affine transformation between V and U and then calculating their bending energy as the final shape-correspondence error. Then, for the correspondence results from different clip lengths, we pick the one with the smallest such normalized

correspondence error as the desirable correspondence. In the next section, we will show that, given the flexility of the landmark-sliding algorithm, the desirable correspondence with smallest normalized correspondence error can be obtained from a certain range of the clip length.

4 Experiments and Analysis

To test the proposed method, we used a fixed template of 41 landmark points and a data set of 32 open-curve femur shapes extracted from medical images. We test the effectiveness of the method by comparing the initial normalized bending energy to the final normalized bending energy (after applying the landmark-sliding algorithm) for each clip length. We can see that the desirable correspondence with minimum normalized bending energy can be obtained from a wide range of clip lengths between 8% and 20%. Therefore, the optimal clip length is not unique and can take a range of values. A sample of the visual results for the optimal clip length can be seen in Fig. 1(c-g) along with (b), which shows the polyline interpolated landmarks in the template shape.

Evaluation and comparison among shape correspondence methods is a difficult problem in which the ground truth is often unknown. As such, we focus on known metrics with which to compare the performance of the proposed method to extant methods for shape correspondence, namely a popular implementation of MDL capable of operating with *nonfixed endpoints* in which the landmarks at the endpoints of the contour are allowed to vary to the same extent as any other landmark in the set[12] and the most recent incarnation of the landmark sliding algorithm [9]. The same template shape was used for all of the methods under consideration to allow for a fair comparison. A brief explanation of the metrics used for comparison follows: (a) $E(\beta)$ and $std(\beta)$: the mean and standard deviation of the thin-plate bending energy between the template and all the shape instances according to the identified landmarks; (b) λ_1, λ_2, and λ_3: the three principal eigenvalues of the covariance matrix of $U_i, i = 1, 2, \ldots, n$. These are the principal zeros of the characteristic polynomial describing the covariance of the data set, which should be as small as possible if the regularity of the shape has been extracted successfully. In calculating the covariance matrix, the Procrustes analysis [1] is applied to normalize the size and orientation of all the shape instances. (c) the sum of the eigenvalues of the covariance matrix of $U_i, i = 1, 2, \ldots, n$; (d) G, the generality of the resulting shape model, calculated according to the leave-one-out cross-validation measure common to statistical processing [11]; and (e) the total CPU time used for processing each data set, based on the adopted specific implementations. In general, with a similar representation error, a good shape correspondence is expected to have small $E(\beta)$, λ_1, λ_2, λ_3, $\sum \lambda$, and G. The statistical improvement gained by this method can be seen in Table 1 as it clearly outperforms the sliding method at a small cost of CPU time and produces results that are comparable to or better than the MDL implementation using free endpoints.

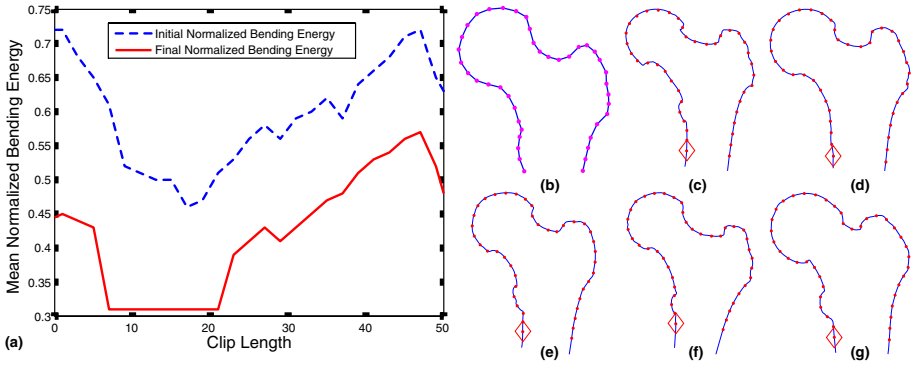


Fig. 1. (a) Comparison results for clip length showing the initial and final normalized bending energy. This bending energy shown here is the mean value on all 32 shape instances. (b) Template shape and (c-g) identified target landmarks on five sample shape instances using the proposed method.

Table 1. Statistical comparison of correspondence results

Measures	MDL [12]	Landmark Sliding developed in [9]	Proposed Method
$E(\beta)$	1.2304	0.4628	0.3149
$std(\beta)$	0.3373	0.1125	0.0824
λ_1	0.0024	0.0030	0.0026
λ_2	0.0008	0.0007	0.0007
λ_3	0.0005	0.0006	0.0005
$\sum \lambda$	0.0051	0.0059	0.0046
$G(\cdot)$	0.8722	0.6337	0.4714
CPU time(s)	418.7	124.7	160.6

5 Conclusion

Because most of the current shape correspondence methods can only effectively deal with complete-shape correspondence, where a one-to-one mapping is assumed between any two shape instances, we present a novel algorithm to correspond open-curve shape instances where a one-to-one correspondence does not exist across all shape instances. In this algorithm, we first introduce an initialization by allowing some portion along the open-curve shape contours to be clipped. We then apply the landmark-sliding algorithm to find the correspondence from each starting clip length. Finally, we introduce a normalized shape-correspondence error to decide the desirable correspondence. The proposed algorithm is tested on 32 femur shape instances and compared to the current landmark-sliding algorithm [9] and an implementation of the MDL method [12]. In this comparison, there is significant statistical improvement using the

proposed algorithm compared to the other current methods with only a small CPU time increase over the current landmark-sliding algorithm.

Acknowledgement

This work was funded, in part, by NSF-EIA-0312861 (S. Wang) and an NSF graduate fellowship (T. Richardson).

References

1. F. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–585, June 1989.
2. F. Bookstein. Landmark methods for forms without landmarks: Morphometrics of group differences in outline shape. *Medical Image Analysis*, 1(3):225–243, 1997.
3. T. Cootes, C. Taylor, D. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, Jan. 1995.
4. R. Davies, C. Twining, T. Cootes, J. Waterton, and C. Taylor. A minimum description length approach to statistical shape modeling. *IEEE Transactions on Medical Imaging*, 21(5):525–537, May 2002.
5. J. Duchon. Splines minimizing rotation-invariant semi-norms in Sobolev space. In *Constructive Theory of Functions of Several Variables, Lecture Notes in Mathematics, 571*, pages 85–100, 1977.
6. T. Heimann, I. Wolf, T. Williams, and H.-P. Meinzer. 3d active shape models using gradient descent optimization of description length. In *Information Processing for Medical Imaging Conference*, 2005.
7. D. Kendall, D. Barden, T. Carne, and H. Le. *Shape and Shape Theory*. John Wiley & Sons, LTD, 1999.
8. M. Leventon, E. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 316–323, 2000.
9. T. Richardson and S. Wang. Nonrigid shape correspondence using landmark sliding, insertion and deletion. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, pages II–435–442, 2005.
10. C. Small. *The Statistical Theory of Shape*. Springer-Verlag, 1996.
11. M. Styner, K. Rajamani, L.-P. Nolte, G. Zsemlye, G. Szekely, C. Taylor, and R. Davies. Evaluation of 3D correspondence methods for model building. In *Information Processing for Medical Imaging Conference*, 2003.
12. H. Thodberg. Minimum description length shape and appearance models. In *IPMI*, pages 51–62, 2003.
13. S. Wang, T. Kubota, and T. Richardson. Shape correspondence through landmark sliding. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages I–143–150, 2004.
14. J. Xie and P. Heng. Shape modeling using automatic landmarking. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, pages II–709–716, 2005.

Reconstruction of Patient-Specific 3D Bone Surface from 2D Calibrated Fluoroscopic Images and Point Distribution Model

Guoyan Zheng[1], Miguel Á.G. Ballester[1], Martin Styner[2], and Lutz-Peter Nolte[1]

[1] MEM Research Center, University of Bern, CH-3014, Bern, Switzerland
Guoyan.Zheng@MEMcenter.unibe.ch
[2] Departments of Computer Science and Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3175, USA

Abstract. Reconstruction of patient-specific 3D bone surface from 2D calibrated fluoroscopic images and a point distribution model is discussed. We present a 2D/3D reconstruction scheme combining statistical extrapolation and regularized shape deformation with an iterative image-to-model correspondence establishing algorithm, and show its application to reconstruct the surface of proximal femur. The image-to-model correspondence is established using a non-rigid 2D point matching process, which iteratively uses a symmetric injective nearest-neighbor mapping operator and 2D thin-plate splines based deformation to find a fraction of best matched 2D point pairs between features detected from the fluoroscopic images and those extracted from the 3D model. The obtained 2D point pairs are then used to set up a set of 3D point pairs such that we turn a 2D/3D reconstruction problem to a 3D/3D one. We designed and conducted experiments on 11 cadaveric femurs to validate the present reconstruction scheme. An average mean reconstruction error of 1.2 mm was found when two fluoroscopic images were used for each bone. It decreased to 1.0 mm when three fluoroscopic images were used.

Keywords: point distribution model, surface reconstruction, 2D/3D correspondence, extrapolation, deformation, thin-plate splines.

1 Introduction

Constructing three-dimensional (3D) bone surface from a few two-dimensional (2D) calibrated fluoroscopic images is a challenging task. *A priori* information is often required to handle this otherwise ill-posed problem. In Fleute and Lavallée [1], a point distribution model (PDM) of distal femur was iteratively fitted to the bone contours segmented on the X-ray views by sequentially optimizing the rigid and non-rigid parameters. It utilizes the principle of the shortest distance between the projection ray of an image edge pixel and a line segment on the apparent contour to set up image-to-model correspondence. However, this method requires an explicit contour extraction, which may be quite difficult when the shapes involved become complex or when the background of the images becomes complex. Furthermore, least-squares based fit may result in unstable solution.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 25–32, 2006.

In Benameur et al. [2][3], a PDM of scoliotic vertebrae was fitted to two conventional radiographic views by simultaneously optimizing both shape and pose parameters. The optimal estimation was obtained by iteratively minimizing a combined energy function, which is the sum of a likelihood energy term measured from an edge potential field on the images and a prior energy term measured from the statistical shape models. No explicit image-to-model correspondence was used. To avoid trapping in a local minimal, their method requires a close initialization.

This paper presents a 2D/3D reconstruction scheme combining statistical extrapolation and regularized shape deformation with an iterative image-to-model correspondence establishing algorithm, and shows its application to reconstruct the surface of proximal femur. The image-to-model correspondence is established using a non-rigid 2D point matching process, which iteratively uses a symmetric injective nearest-neighbor mapping operator and 2D thin-plate spline (TPS) based deformation to find a fraction of best matched 2D point pairs between features detected from the fluoroscopic images and those extracted from the 3D model. The obtained 2D point pairs are then used to set up a set of 3D point pairs such that we turn a 2D/3D reconstruction problem to a 3D/3D one. No explicit contour extraction from fluoroscopic images is required in the present scheme.

This paper is organized as follows. Section 2 presents the point distribution model. Section 3 describes the iterative image-to-model correspondence establishing algorithm. Section 4 briefly recalls the 3D/3D reconstruction framework introduced in [4]. Section 5 presents the experimental results, followed by the conclusions in Section 6.

2 Point Distribution Model

The PDM used in this paper was constructed from a training database consisted of 30 proximal femoral surfaces from above the less trochanter. To optimally align these training surfaces, a sequence of correspondence establishing methods presented in [5] was employed. It started with a SPHARM-based parametric surface description [6] and then was optimized using Minimum Description Length (MDL) based principle as proposed by Davis et al [7].

Following the alignment, the PDM is constructed as follows. Let $\mathrm{x}_i, i = 0, 1, ..., m-1$, be m (here m=30) members of the aligned training surfaces. Each member is described by a vectors x_i with N vertices:

$$\mathrm{x}_i = \{x_0, y_0, z_0, x_1, y_1, z_1, ..., x_{N-1}, y_{N-1}, z_{N-1}\} \tag{1}$$

The PDM is obtained by applying principal component analysis.

$$\begin{array}{l} D = ((m-1)^{-1}) \cdot \sum_{i=0}^{m-1} (\mathrm{x}_i - \bar{\mathrm{x}})(\mathrm{x}_i - \bar{\mathrm{x}})^T \\ P = (\mathbf{p}_0, \mathbf{p}_1, ...); D \cdot \mathbf{p}_i = \sigma_i^2 \cdot \mathbf{p}_i \end{array} \tag{2}$$

where $\bar{\mathrm{x}}$ and D are the mean vector and the covariance matrix, respectively.

Then, any one of the instance in this space can be expressed as:

$$\mathrm{x} = \bar{\mathrm{x}} + \sum_{i=0}^{m-2} \alpha_i \mathbf{p}_i \tag{3}$$

And the estimated normal distribution of the coefficients α_i is:

$$p(\alpha_0, \alpha_1, ..., \alpha_{m-2}) = (2\pi)^{-\frac{m-1}{2}} \cdot \exp(-\frac{1}{2} \sum_{i=0}^{m-2} (\alpha_i^2/\sigma_i^2)) \tag{4}$$

where $\sum_{i=0}^{m-2} (\alpha_i^2/\sigma_i^2)$ is the Mahalanobis distance defined on the distribution.

3 Image-to-Model Correspondence

Given a few fluoroscopic images, our task is to establish correspondence between the input images and a model estimated from the PDM. Here we assume that the input images are calibrated and registered to a common coordinate system. And for a pixel in an input image we can always find a projection ray emitting from the focal point of the associated image through the pixel.

A. Edge detection: A standard implementation of Canny edge detector with hysteresis [8] is used to find the edge pixels of the considered bone structure from the input images. The detected edge pixels are further processed using the knowledge about the X-ray machine. Detected edge pixel whose intensity is below some threshold or whose position is of a certain distance away from the image center is eliminated.

B. Apparent contour extraction: For fast extraction a smooth apparent contour from a given surface model, we use the approach introduced by Hertzmann and Zorin [9]. This approach first constructs a dual representation of the given surface mesh in a four-dimensional (4D) space based on the position and tangent planes of every vertex. The focal point's dual (a plane in 4D) intersects with the mesh triangles dual. Before hand, the approach normalizes the dual vertices using the l_∞ norm so that the vertices end up on one of the unit hyper-cube's sides. This reduces the problem to intersecting the triangles on a hypercube's sides with the focal point's dual plane, whose solution can be accelerated greatly using octree data structure. An example is given in Fig. 1.

Then, for a point $\Omega_j^s, (j = 0, 1, ..., M-1)$ on the extracted apparent contours using the projection parameters of the sth input X-ray image, we do a forward projection of this point onto image s to get its 2D position A_j^s. Each 2D point in $\{A_j^s\}$ is thus associated to a 3D point in $\{\Omega_j^s\}$. Next, we will describe an iterative matching process for build 2D association between the points in $\{A_j^s\}$ and the detected edge pixels in the X-ray image s.

C. Iterative non-rigid 2D matching process: Following the general framework of point matching proposed in [10], we also formulate the 2D matching as a two-stage process: correspondence and spatial mapping. TPS is used here for parameterization of the spatial mapping. But unlike [10], we solve the correspondence problem differently. To make the description simple, we denote

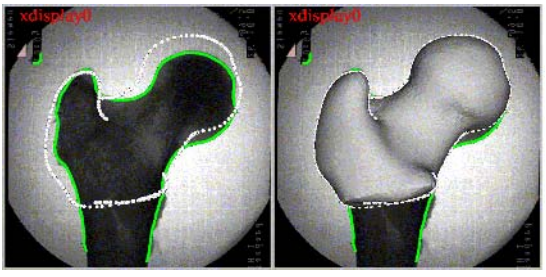


Fig. 1. Example of extracting apparent contours (white) and edge pixels (green)

the detected edge pixels in image s as $I^s = \{I_i^s, i = 0, 1, ..., Q-1\}$ and the 2D projection of the apparent contours as $A^s = \{A_j^s, j = 0, 1, ..., M-1\}$. Here we focus on 2D matching in one image. The overall correspondence is established by combining matched point pairs found in all input images.

Definition 1: Injective Nearest-neighbor. A point I_i^s can only be matched to at most one point of A^s and this point must be the closest one in A^s to I_i^s.

Definition 2: Cross-matching. Assume there are two matched pairs (I_i^s, A_j^s) and (I_m^s, A_n^s). If the line segmentation from I_i^s to A_j^s intersects at a point with the line segment from I_m^s to A_n^s, we define this event as cross-matching.

Definition 3: Symmetric Injective Nearest-neighbor (SIN). I_i^s and A_j^s is a symmetric injective nearest-neighbor if and only if A_j^s is the closest point in A^s to I_i^s and I_i^s is the closest point in I^s to A_j^s.

Definition 4: Symmetric Injective Nearest-neighbor Mapping Operator (SIN-MO). We define the action of finding a number of SIN's from two points set as symmetric inject nearest-neighbor mapping operator.

Claim 1: If we apply SIN-MO on two point sets to find a number of matched point pairs, all of them are one-to-one mapping (proof omitted).

Claim 2: If we apply SIN-MO on two point sets to find a number of matched point pairs, there is no cross-matching (proof omitted).

The iterative non-rigid 2D matching process can be described as follows.

Input: two point sets I^s and A^s, and a weight parameter λ.

Output: a list of 2D matched point pairs.

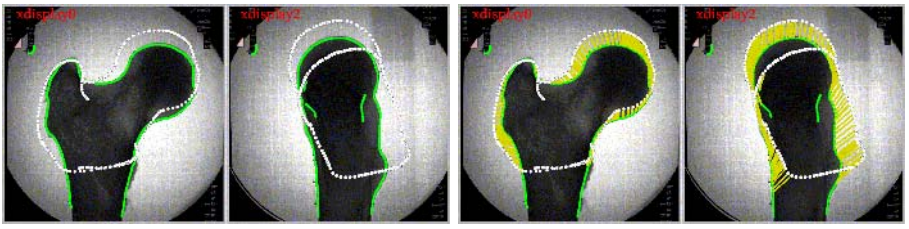
Initialization: We first calculate the centroids of I^s and A^s, and then translate A^s so that its centroid is aligned with the centroid of I^s.

Iteration: It is a dual update process taking points in I^s as references.

Stage 1: Update the correspondence. Apply SIN-MO on I^s and A^s to find a number of SINs and denote the set of SINs as $\{(I_a^s, A_a^s); a = 0, 1, ..., K-1\}$.

Stage 2: Update the positions of all points in A^s. This is done in a two-step procedure.

Step 2.1. Compute a 2D TPS-based spatial mapping f using the estimated set of SINs by minimizing the following cost function:



(a) Before building correspondence (b) After building correspondences

Fig. 2. Establishing 2D/3D correspondence; detected edge pixels (green), extracted apparent contours (white), and the estabilished correspondences (linked with yellow line segment for visualization purpose)

$$\sum_{a=0}^{K-1} ||I_a^s - f(A_a^s)||^2 + \lambda \iint [(\frac{\partial^2 f}{\partial x^2})^2 + (\frac{\partial^2 f}{\partial x \partial y})^2 + (\frac{\partial^2 f}{\partial x^2})^2] dxdy \qquad (5)$$

Step 2.2. Update the positions of all points in A^s based on the estimated TPS transformation f.

Repeat stage 1 and 2 a certain times (e.g. 30) or until convergence.

D. 3D point pair building: Assume that we have found a set of 2D matched point pairs $\{(I_b^s, A_b^s); b = 0, 1, ..., L-1\}$, we are trying to build the corresponding 3D point pairs as follows. For a 2D point I_b^s, we can find a projection ray r_b^s emitting from the focal point of image s through point I_b^s. Additionally, for its matched point A_b^s, we always know its associated 3D point Ω_b^s on the apparent contour of the model whose projection onto the image s is A_b^s. By computing a point v_b^s on the ray r_b^s that has the shortest distance to Ω_b^s, we can build a 3D point pair (v_b^s, Ω_b^s). Combining all these 3D point pairs, we can establish 2D/3D correspondence between the input images and a 3D model estimated from the PDM. An example is given in Fig. 2.

4 3D/3D Reconstruction

Using the developed algorithm for establishing correspondence, we can always find a set of 3D point pairs given an initial model state, i.e., the initial scale and the initial pose parameters of our point distribution model. The problem of surface reconstruction is then solved optimally in three stages as presented in our previous work [4]: *affine registration*, *statistical extrapolation*, and *regularized shape deformation*.

A. Affine registration: This is the only stage that is solved by iteration. In this stage, the scale and the rigid registration transformation between the mean model of the PDM and the input images are iteratively determined using a generalization of the Iterative Closest Point (ICP) algorithm developed by Guéziec et al. [11]. The difference between this algorithm and the traditional ICP algorithm is that in each iteration we need to set up a set of new point pairs using the present algorithm for building 2D/3D correspondences.

B. Statistical extrapolation: Based on the estimated scale and pose information from the first stage, we can use the correspondence establishing algorithm to obtain a set of n 3D point pairs from all input images. Let's denote the points computed from the image data as $\mathrm{v}' = \{v_i' = (x_i', y_i', z_i'); i = 0, 1, ..., n-1\}$ and call them the image points. Let's further denote those points on the mean model of the PDM as $\bar{\mathrm{x}}' = \{(\bar{\mathrm{x}}_j)_i; 0 \le j \le N-1; i = 0, 1, ..., n-1\}$ and call them the model points, where N is the number of points of the mean model; n is the number of the matched point pairs. $(\bar{\mathrm{x}}_j)_i$ means that the jth model point $\bar{\mathrm{x}}_j$ on the mean model $\bar{\mathrm{x}}$ is the closest point to the ith image point v_i'. The statistical extrapolation is formulated as the minimization of the following joint cost function:

$$\begin{aligned} &E_\alpha(\bar{\mathrm{x}}', \mathrm{v}', \mathrm{x}) = \rho \cdot \log(3n) \cdot E(\bar{\mathrm{x}}', \mathrm{v}', \mathrm{x}) + \mathrm{E}(\mathrm{x}); \mathrm{x} = \bar{\mathrm{x}} + \textstyle\sum_{k=0}^{m-2} \alpha_k \cdot \mathbf{p}_k \\ &E(\bar{\mathrm{x}}', \mathrm{v}', \mathrm{x}) = (n^{-1}) \cdot \textstyle\sum_{i=0}^{n-1} ||v_i' - ((\bar{\mathrm{x}}_j)_i + \sum_{k=0}^{m-2} \alpha_k \cdot \mathbf{p}_k(j))||^2 \\ &E(\mathrm{x}) = (1/2) \cdot \textstyle\sum_{k=0}^{m-2} (\alpha_k^2 / \sigma_k^2) \end{aligned} \tag{6}$$

where the first term is the likelihood energy term and the second term is the prior energy term (or the stabilization term), used to constrain the estimated shape to a realistic result. ρ is a parameter that controls the relative weighting between these two terms. $\mathbf{p}_k(j)$ is the jth tuple of the kth shape basis eigenvector.

For details about how to solve Eq. (6), we refer to our previous work [4].

C. Regularized shape deformation: Similar to the second stage, we first set up image-to-model correspondence between the input images and a template surface model estimated through optimizing Eq. (6). To keep the same notation, let's assume that the image points are $\mathrm{v}' = \{v_i' = (x_i', y_i', z_i'); i = 0, 1, ..., l-1\}$ and that the model points are $\mathrm{v} = \{v_i = (\mathrm{x}_j)_i = (x_i, y_i, z_i); i = 0, 1, ..., l-1\}$, where l is the number of point pairs and $(\mathrm{x}_j)_i$ means that the jth model point x_j on the statistically extrapolated surface x is the closest point to the ith image point v_i'. The regularized shape deformation is described as a regression problem of finding a spatial transform $\mathrm{t}: \Re^3 \rightarrow \Re^3$ that minimizes following cost function:

$$E(\mathrm{t}) = (l^{-1}) \cdot \sum_{i=0}^{l-1} ||v_i' - \mathrm{t}(v_i)||^2 + \tau \cdot \frac{\log(m)}{\log(3l)} \cdot L[\mathrm{t}] \tag{7}$$

where $\tau \ge 0$ is a parameter controlling the fitting quality and the regularization constraint; m is the number of training surfaces; $\mathrm{t}(\mathrm{v}) = \{\mathrm{t}(v_i)\}$ are the results of applying the mapping on the model points and $L[\mathrm{t}]$ is a thin-plate splines based regularization functional defined on the nonlinear mapping t and has following form [12]:

$$\begin{aligned} &L[\mathrm{t}] = \iiint_{\Re^3} (B(\mathrm{t}))dxdydz \\ &B(\cdot) = (\frac{\partial^2}{\partial x^2})^2 + (\frac{\partial^2}{\partial x \partial y})^2 + (\frac{\partial^2}{\partial y^2})^2 + (\frac{\partial^2}{\partial y \partial z})^2 + (\frac{\partial^2}{\partial z^2})^2 + (\frac{\partial^2}{\partial z \partial x})^2 \end{aligned} \tag{8}$$

For details about how to solve Eq. (7), we refer to our previous work [4].

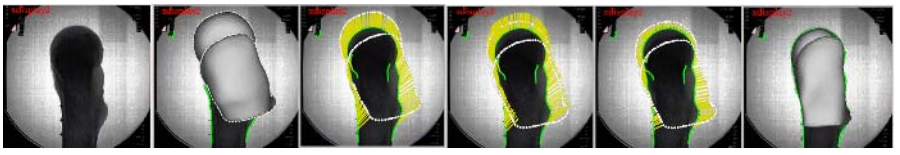


Fig. 3. Different stages of reconstruction. First: one of the acquired images. Second: the initialization of the mean model of the PDM. Third: after establishing image-to-model correspondence. Forth: after 3D paired point matching. Fifth: after re-establishing correspondence; Sixth: the final reconstruction result after a series of computations

Table 1. Reconstruction errors when different number of images were used

Reconstruction errors when only AP and LM images were used for each bone											
Bone Index	No. 1	No. 2	No. 3	No. 4	No. 5	No. 6	No. 7	No. 8	No. 9	No. 10	No. 11
Median (mm)	1.3	0.8	1.5	1.0	1.3	1.0	1.1	1.0	0.8	1.1	1.2
Mean (mm)	1.5	0.8	1.4	1.3	1.4	1.2	1.2	1.2	1.0	1.1	1.6
Reconstruction errors when all three images were used for each bone											
Bone Index	No. 1	No. 2	No. 3	No. 4	No. 5	No. 6	No. 7	No. 8	No. 9	No. 10	No. 11
Median (mm)	1.3	0.7	0.7	1.1	1.0	1.1	0.8	0.9	0.7	1.0	0.9
Mean (mm)	1.3	0.7	0.8	1.2	1.1	1.1	1.1	0.9	0.9	1.1	1.2

5 Experiments and Results

We designed and conducted experiments on 11 cadaveric femurs (Note: none of them has been included for constructing the PDM) with different shape to validate the present reconstruction scheme. Three fluoroscopic images were acquired for each bone: the first one along the anterior-posterior (AP) direction, the second one along the lateral-medial (LM) direction, and the last one from an oblique angle in between the AP and the LM directions. All acquired images were calibrated and registered to the associated reference coordinate system. And in all reconstructions, three controlling parameters are set as follows: λ=0.5, ρ=0.05, and τ= 0.05.

Two experiments using different number of images were performed for each bone. In the first experiment only AP and LM images were used to reconstruct the surface of the bone. In the second one, all three images were used.

The reconstruction accuracies were evaluated by randomly digitizing 100 - 200 points from each surface of the cadaveric specimen and then computing the distance from those digitized points to the associated surface reconstructed from the images. The median and mean reconstruction errors of both experiments are presented in Table 1. An average mean reconstruction error of 1.2 mm was found when only AP and LM images were used for each bone. It decreased to 1.0 mm when three images were used. Different stages of one reconstruction example are presented in Fig. 3.

6 Conclusions

In this paper, we have presented a 2D/3D reconstruction scheme combining statistical extrapolation and regularized shape deformation with an iterative image-to-model correspondence establishing algorithm, and showed its application to reconstruct the surface of proximal femur. The advantages of the present image-to-model correspondence establishing algorithm include the robustness with respect to certain outliers and automatic exclusion of cross matching, which is an important property for preservation of topology. Experiments performed on 11 cadaveric femurs demonstrated the accuracies of the present scheme. It holds the potential to increase reconstruction accuracy when more images are used. Our future plan is to improve the algorithm for image feature detection with the final goal of applying the present reconstruction scheme to real situation.

References

1. Fleute M., Lavallée S.: Nonrigid 3D/2D registration of images using a statistical model. Lecture Notes in Computer Science, Springer **1679** (1999) 138–147
2. Benameur S., Mignotte M., Parent S. et al.: 3D/2D registration and segmentation of scoliotic vertebra using statistical models. Comput Med Imag Grap **27** (2003) 321–337
3. Benameur S., Mignotte M., Parent S., et al.: A hierarchical statistical modeling approach for the unsupervised 3D reconstruction of the scoliotic spine. In ICIP'04 (2004) 561–563
4. Zheng G., Rajamani K.T. and Nolte L.-P.: Use of a dense surface point distribution model in a three-stage anatomical shape reconstruction from sparse information for computer-assisted orthopaedic surgery: a preliminary study. In ACCV'06, Lecture Notes in Computer Science, Springer **3852** (2006) 52–60
5. Styner M., Rajamani K.T., Nolte L.-P., Zsemlye G., Szkely G., Taylor C.J., Davis R.H.: Evaluation of 3D correspondence methods for modeling building. In IPMI'03, Lecture Notes in Computer Science, Springer **2732** (2003) 63–75
6. Brechbuehler C., Gerig G., Kuebler O.: Parameterization of Closed Surfaces for 3D Shape Description. Comput Vision and Image Under **61** (1995) 154–170
7. Davies R. H., Twining C. H., et al.: 3D statistical shape models using direct optimization of description length. In ECCV'02, Lecture Notes in Computer Science, Springer **2352** (2002) 3–20
8. Canny J.: A computational approach to edge detection. IEEE T Pattern Anal **8** (1986) 679–698
9. Hertzmann A., Zorin D.: Illustrating smooth surfaces. In SIGGRAPH'00, (2000), 517–526
10. Chui H., Rambo J., Duncan J., Schultz R., and Rangarajan A.: Registration of cortical anatomical structures via robust 3D point matching. In: IPMI'99, Lecture Notes in Computer Science, Springer **1613** (1999), 168–181
11. Guéziec A., Kazanzides P., Williamson B., and Taylor R.H.: Anatomy-based registration of CT-scan and intraoperative X-ray images for guiding a surgical robot. IEEE T Med Imaging, **17** (1998), 715–728
12. Bookstein F.: Principal warps: thin-plate splines and the decomposition of deformations. IEEE T Pattern Anal, **11** (1989), 567–585

A Pilot Study of Robot-Assisted Cochlear Implant Surgery Using Steerable Electrode Arrays

Jian Zhang[1], Kai Xu[1], Nabil Simaan[1], and Spiros Manolidis[2]

Columbia University
[1] Department of Mechanical Engineering
Laboratory of Advanced Robotics and Mechanism Applications
RM 234 S.W. Mudd Building, 500 West 120th Street, New York, NY 10027, USA
{jz2182, kx2001, ns2236}*@columbia.edu
http://www.columbia.edu/cu/mece/arma/
[2] Department of Otolaryngology – Head & Neck Surgery
HP813, 180 Fort Washington Avenue, New York, NY 10032, USA
sm2307*@columbia.edu

Abstract. This paper presents results of a pilot study evaluating the efficacy of robotic assistance using novel steerable electrode arrays for cochlear implant surgery. The current surgical setup of cochlear implant surgery is briefly reviewed and its limitations are highlighted. In an effort to reduce trauma to the structure of the cochlea, the kinematics and path planning for novel cochlear steerable electrodes are developed to minimize the interaction forces between the electrode and the cochlea. An experimental robotic system is used to compare the electrode insertion forces of steerable implants with those of non-steerable electrodes. The results of these experiments show about 70% reduction in the insertion forces when steerable electrodes are used with our proposed path planning and control. A distance metric explaining this reduction in the insertion force is defined and experimentally validated. Although this is only a preliminary study, we believe that these results provide a strong indication to the potential of robot-assisted cochlear implant surgery to provide a significant reduction in trauma rates during cochlear implant surgery.

1 Introduction

Cochlear implant surgery allows surgeons to restore partial hearing to patients suffering from severe hearing loss due to damaged neuroepithelial (hair) cells. During this surgery surgeons insert long, thin, and flimsy electrodes into the scala tympani canal. The interaction forces during this surgery are small (less than 10 grams) [1]. The tools used by surgeons do not provide any force feedback and are often very difficult to control. The complex anatomy of the cochlea does not lend itself to intra-operative imaging of its fine structures such as the basilar membrane. The basilar membrane is very delicate and can be easily ruptured by the inserted implant electrode. All these characteristics of cochlear implant surgery currently limit both its success (lower than 60% of atraumatic insertions [2-4]) and applicability.

* Corresponding authors.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 33 – 40, 2006.

The success of an implanted electrode in restoring a wider frequency range depends on the depth of its atraumatic insertion. While increased depth of insertion promises improvement, it also increases the risk of trauma to the delicate structures of the cochlea. For these reasons cochlear implant surgery is currently characterized by shallow atraumatic insertions (less than 450°). Due to the traumatic nature of the electrode insertions cochlear implant surgery is currently not applicable to patients with residual hearing, but suffering from hearing loss that can not be restored by external hearing aids.

Following previous works on snake-like robots for distal dexterity enhancement [5, 6], this paper quantifies the potential improvement in cochlear implant surgery if robotic-assistance is used with novel steerable electrodes that provide some dexterity inside the cochlea. The design, kinematic modeling, calibration, and path planning of under-actuated flexible steerable electrodes are treated in this paper with a goal of minimizing the interaction forces with the walls of the cochlea. Mathematical modeling and path planning are presented in section 2. Although we present results using novel steerable electrodes currently being developed in our lab, our methodology in section 2 is valid for robotic assisted insertions using current electrodes with off-stylet insertion tools [3]. These tools provide some control over the shape of the electrode while sacrificing force feedback due to the very small interaction forces and friction in the traditional insertion tools. Section 3 presents our experimental results comparing the insertion forces of straight non-steerable electrodes to steerable electrodes and quantifying the benefits of steerable implants.

2 Problem Statement and Mathematical Modeling

The purpose of this preliminary study is to evaluate the efficacy of robot-assisted electrode insertion using novel steerable cochlear implant electrodes. We hypothesize that a reduction in the electrode insertion forces will result in a significant reduction in the trauma rates during cochlear implant surgery[1]. We thus use the insertion force of the electrode as a performance measure to quantify the potential benefits of using these steerable electrodes for reducing trauma in cochlear implant surgery.

Due to the small size of the cochlear implant electrodes we assume that controlling the shape of the electrode will be limited for a single actuator. Accordingly, we treat the steerable electrodes as continuum under-actuated snake-like robots. For these robots the solution of the direct kinematics is determined by the variational solution minimizing the elastic energy of the electrode [7, 8].

Fig. 1 shows scaled up (3:1) steerable electrode models that we fabricated using silicone rubber. These implant models are actuated by a ∅100 μm Kevlar thread. Figure 1-(b) and Figure 1-(c) show two different electrodes that we molded to yield different bending shapes. The problem at hand is to design an electrode that bends into the desired shape that best approximates the shape of the cochlea and to determine the optimal path planning that will provide the best approximation of the shape of the cochlea during the electrode insertion process. This in turn will yield minimal interaction forces with the walls of the cochlea.

[1] This hypothesis will be clinically tested in future studies.

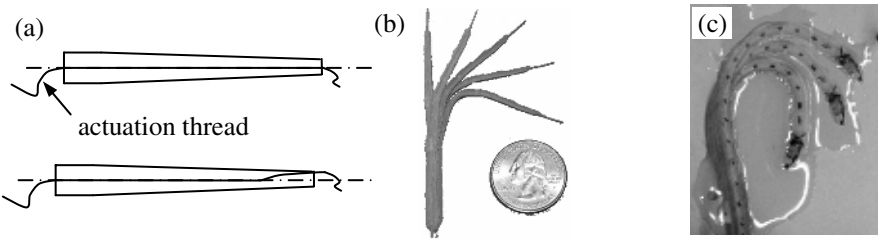


Fig. 1. (a) Top and front view of the electrode, (b-c) Two electrode models with different bending characteristics

Direct kinematics of the steerable electrodes: Let θ and s represent the angle and the arc length along the backbone of the electrode. Let s=0 indicate the base and s=L the tip of the electrode. Let q be the value of the active joint controlling the bending of the electrode. The shape of a planar bending electrode is given by θ(s). Let the minimum energy solution for the direct kinematics of the electrode be approximated using a modal representation [9-12], Eq. (1) where vector **a** is a vector of modal factors. Let this vector of modal factors be given by $\mathbf{a}(q)=\mathbf{A}\boldsymbol{\eta}(q)$, $\boldsymbol{\eta}\in\Re^{m},\mathbf{A}\in\Re^{n\times m}$ where $\boldsymbol{\eta}^{\,t}=[1,q,q^{2},\cdots,q^{m-1}]$. We note that for high-order polynomial approximations (m>6), a set of orthogonal polynomials (e.g. Chebyshev polynomials) should be used for considerations of numerical stability [13].

$$\theta(s,q)=\boldsymbol{\psi}(s)^{t}\mathbf{a}(q),\quad \mathbf{a},\boldsymbol{\psi}\in\Re^{n}\ \text{where}\ \boldsymbol{\psi}(s)=\left[1,s,\ldots,s^{n-1}\right]^{t} \tag{1}$$

Since the minimum-energy solution does not lend itself to real-time control we choose to calibrate the direct kinematics experimentally. The shape of the electrode may be digitized by r equidistant points along its backbone in z different images of the electrode associated with z different values of q. For each point along the backbone the angle of the curve tangent is digitized and recorded in an experimental data matrix $\boldsymbol{\Phi}\in\Re^{r\times z}$ such that $\Phi_{i,j}=\theta(s_i,q_j)$. Using the modal representation in Eq. (1) the direct kinematics problem is cast as an algebraic matrix equation, Eq. (2).

$$\boldsymbol{\Phi}=\underbrace{\begin{bmatrix}\boldsymbol{\psi}^{t}(s=0)\\ \vdots\\ \boldsymbol{\psi}^{t}(s=L)\end{bmatrix}}_{\boldsymbol{\Omega}}\mathbf{A}_{n\times m}\underbrace{\left[\boldsymbol{\eta}(q=0),\ \cdots,\ \boldsymbol{\eta}(q=q_{max})\right]}_{\boldsymbol{\Gamma}}=\boldsymbol{\Omega}_{r\times n}\mathbf{A}_{n\times m}\boldsymbol{\Gamma}_{m\times z}$$

$$\boldsymbol{\Omega}=\begin{bmatrix}1 & s & \cdots & s^{n-1}\\ & \vdots & & \\ 1 & s & \cdots & s^{n-1}\end{bmatrix}_{r\times n}\qquad \boldsymbol{\Gamma}=\begin{bmatrix}1 & 1 & & 1\\ q & q & \cdots & q\\ \vdots & \vdots & & \vdots\\ q^{m-1} & q^{m-1} & & q^{m-1}\end{bmatrix}_{m\times z} \tag{2}$$

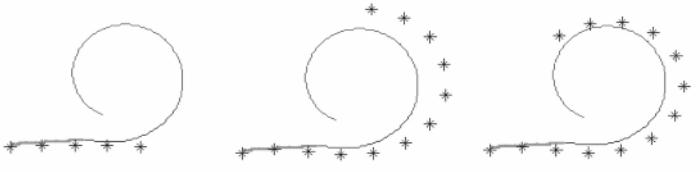
$\boldsymbol{\Omega}$ and $\boldsymbol{\Gamma}$ are Vandermonde matrices corresponding to the r numerical values of s and the z values of q used to generate the experimental data matrix $\boldsymbol{\Phi}$. Solving Eq. (2) ($\boldsymbol{\Phi}=\boldsymbol{\Omega}\mathbf{A}\boldsymbol{\Gamma}$) for matrix **A** provides the required solution for the direct kinematics

problem. The solution of this algebraic matrix equation is given by $[\,\mathbf{\Gamma}^{t} \otimes \mathbf{\Omega}]\,\mathrm{Vec}(\mathbf{A}) = \mathrm{Vec}(\mathbf{\Phi})$ [14]. Where $\otimes$ represents Kronecker's matrix product and $Vec(\mathbf{A}_{m\times n}) = [a_{11}\cdots a_{m1}, a_{12}\cdots a_{m2}, a_{1n}\cdots a_{mn}]^{t}$.

Optimal electrode insertion path planning: Let s_q represent the electrode insertion depth and let $\theta_c(s)$ be the shape of the cochlea. Equation (3) returns the optimal value of q that minimizes the shape difference between the inserted portion of the electrode and the cochlea. The optimal value of q is found by calculating the objective function for all columns of $\mathbf{\Phi}$ and the minimum is found by numerical interpolation between the columns that best approximate the minimum value of the objective function.

$$\arg\min_{q} \int_{L-s_q}^{L} (\left|\theta_c(s) - \theta(s)\right|^2) \tag{3}$$

Fig. 2 shows the results of a Matlab simulation of this path planning process applied using the experimental data matrix $\mathbf{\Phi}$ of the electrode of Fig. 1-(c). The figure demonstrates that the implant coincides with the cochlear model well except for the tip because the implant does not actively bend at its tip.



✳ = Indicates the shape of steerable electrode.
Solid line = shape of the cochlea using 2D Cohen's template up to 330° insertion angle.

Fig. 2. Path planning insertion simulation for the electrode in Fig. 1-(c)

3 Experimental Validation

Phantom models of the cochlea: The cochlea has a 3-dimensional spiral anatomical structure that was statistically characterized in the work of Cohen, et. al. [15], and extended to 3D by Ketten et. al. [16]. The backbone curve of the cochlea is given by Eq. (4) where *r, z,* and θ are the cylindrical coordinates of this curve (*r* is the radial distance to the curve, *z* is the height, and θ is the angle). The values of the constants *a, c, b, d,* θ_0, *p* are based on [16], [17].

Fig. 3-(a-b) show a CAD model of the cochlea and the scala tympani. Fig. 3-(c) shows a 3-dimensional stereo lithography 2:1 model of scala tympani. Figure Fig. 3-(d) shows a 2-dimensional 3:1 scaled up model of the cochlea using Cohen's 2D template. The 2-dimensional phantom model in Fig. 3-(d) was used in this work for electrode insertion evaluation. Although this model does not provide insertion

angles larger than 340°, it is enough to demonstrate the differences between passive electrodes and active bending electrodes used in our work.

$$\theta \in \left[10.3^\circ, 910^\circ\right], \quad R = \begin{Bmatrix} c\left(1 - d\log\left(\theta - \theta_0\right)\right) & \theta < 100^\circ \\ ae^{-b\theta} & \theta \geq 100^\circ \end{Bmatrix}, \quad z = p\left(\theta - \theta_0\right) \tag{4}$$

A two Degrees-of-Freedom (DoF) experimental robotic insertion system with force monitoring: The experimental system shown in Fig. 4 was used to compare electrode insertion forces for steerable and non-steerable electrodes. This system is composed from a single DoF linear actuator, AG NTEP 5000d single axis load cell, and a steerable implant that is inserted into the 3:1 phantom model of Figure 3-(d). The electrode model is supported against buckling using an implant support ring that is affixed to the force sensor. This support ring was placed in two axial positions: one for shallow insertions, and another for deep insertions, Fig. 5. Since the direct kinematics of the implant is affected by the constraint imposed by the slip ring, we calibrated the direct kinematics for both configurations in Fig. 5.

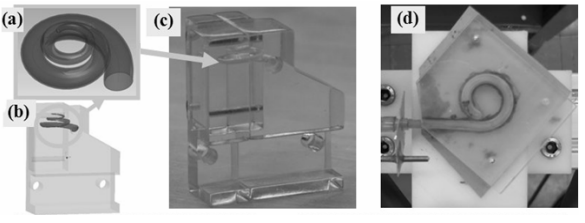


Fig. 3. (a) A 3D CAD model of the cochlea. (b) A CAD model of the scala tympani. (c) A 2:1 3D stereolithography model of the scala tympani. (d) A planar 3:1 model of the cochlea.

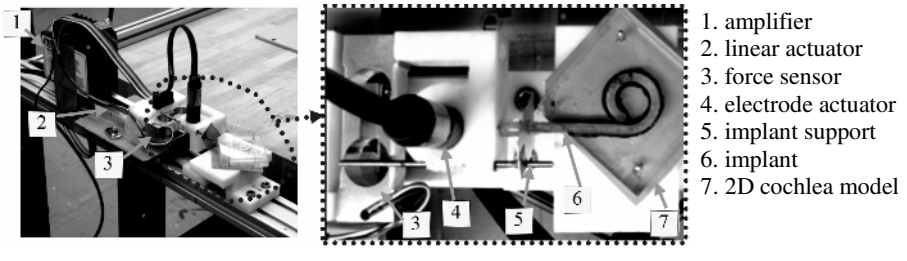


Fig. 4. Experimental system setup for robot-assisted electrode insertion

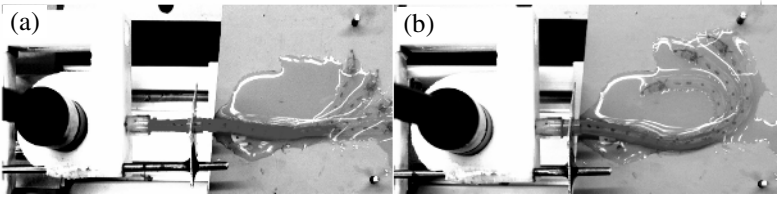


Fig. 5. Two configurations of the implant buckling support ring were used: (a) Configuration for shallow insertion. (b) Configuration for deep insertion.

Direct kinematics calibration: For the electrode direct kinematics calibration we marked the electrode with sixteen equidistant points along its backbone. Twelve images with different total bending angles of the electrode were acquired and the position of the implant support ring was changed from a pre-set value for shallow insertions to another pre-set value for deep insertions, Fig. 5.

Insertion experiments: The same electrode was inserted three times without controlling its bending (the electrode was set in a straight shape outside of the cochlea). Another set of three insertions was performed while controlling the bending of the electrode based on the path planning algorithm of section 2. As in [1], we wetted the cochlea with glycerin to emulate the friction conditions during surgery. The forces of insertion were recorded as a function of the insertion depth of the electrode.

Results: Fig. 6-(a) shows the average insertion forces over three insertion experiments using the electrode of Fig. 1-(c) with and without control of its bending. Fig. 6-(b) compares the best insertion of the electrode without bending control with the three insertion experiments using the path planning of section 3. The results show that the active bending insertions are repeatable and that the best non-steerable insertion is significantly worse than the worst steerable insertion. A prominent reduction of **68%** in the insertion force was achieved when using our path planning.

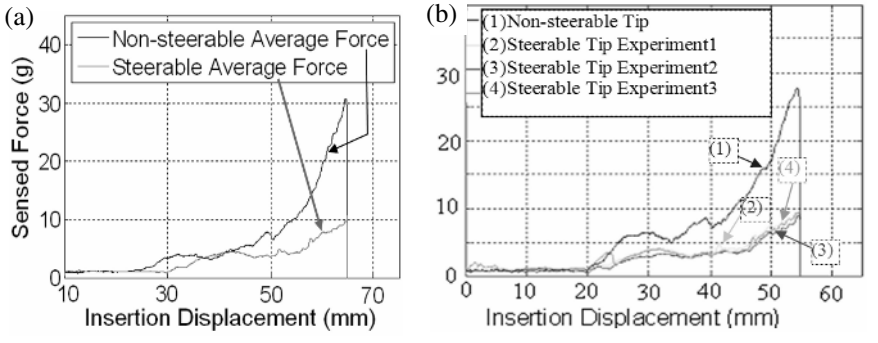


Fig. 6. (a) Average insertion forces of steerable vs. non-steerable electrode over three insertions. (b) The insertion forces for all the steerable electrode insertions are significantly lower than the best non-steerable implant insertion.

Analyzing experimental results: The insertion force due to friction between the electrode and cochlea is equivalent to friction force in a band brake system [18], which depends on the contact angle of the electrode with the external walls of the cochlea. To explain the results of Fig. 6, several images during the insertion experiments were acquired for several insertion depths and digitized as shown in Fig. 7-(a). Third-order polynomials were fitted to the digitized data to represent the curve of the external wall of the electrode, $\mathbf{r}_c$, and the curve of the outer wall of the cochlea , $\mathbf{r}_I$, Fig. 7-(b). Using these polynomial representations a distance metric $e(\theta) = \|\mathbf{r}_c(\theta) - \mathbf{r}_I(\theta)\|_2 \quad \theta \in [0, \varphi]$ was calculated (where φ is the insertion angle) and averaged for every insertion angle during the insertion, Eq. (5).

$$\bar{e} = \varphi^{-1} \int_0^{\varphi} e(\theta)\, d\theta \tag{5}$$

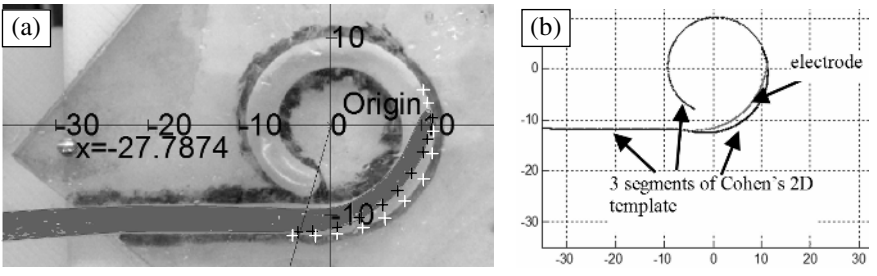


Fig. 7. (a) A sample digitized image. (b) The plots of Cohen's 2D template and the curve of the outside wall of the electrode (all units in mm).

Fig. 8 shows the average distance metric $\bar{e}$ vs. the electrode insertion depth. Fig. 8 and Fig. 6 explain the decrease in the insertion forces when the electrode is actuated since the average distance metric is increased significantly compared to the non-steerable electrode. Moreover, the difference between steerable and non-steerable electrodes becomes more prominent as the insertion depth increases.

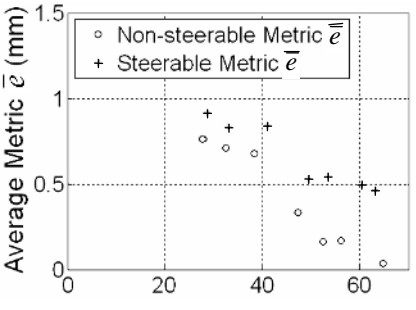


Fig. 8. Average distance metric $\bar{e}$ for Non-steerable and Steerable implant as a function of insertion depth (in mm)

4 Conclusion

This paper presented a pilot study that evaluates the potential benefits of robot-assisted cochlear implant surgery using active-bending electrodes. Although this is a preliminary study using phantom models, it clearly indicates the potential of robot-assisted cochlear implant surgery using steerable electrodes in reducing the trauma to the cochlea. The paper presented the mathematical modeling for the steerable electrodes including a path planning algorithm that minimizes the interaction forces during electrode insertion. An experimental system was presented and the results of insertion experiments comparing insertions using steerable versus non-steerable electrodes showed a reduction of about 70% in the insertion forces when steerable electrodes are used. The paper also correlated the reduction in the insertion forces with an increase in the average distance metric between the electrode and the cochlea as a result of the distal dexterity provided by the proposed steerable electrodes. Based on these results we are currently developing a tele-robotic system for robot-assisted cochlear implant surgery.

References

1. Roland, T., *A model for cochlear implant electrode insertion and force evaluation: results with a new electrode design and insertion technique.* The Laryngeoscope, 2005. **115**(august): p. 1325-1339.
2. Eshraghi, A., N. Yang, and T. Balkany, *Comparative Study of Cochlear Damage with Three Perimodiolar Electride Designs.* The Laryngeoscope, 2003. **113**: p. 415-419.
3. Wardrop, P., et al., *A temporal bone study of insertion trauma and intracochlear position of cochlear implant electrodes I: comparison of Nucleus banded and Nucleus Contour electrodes.* Hearing Research, 2005. **203**: p. 54-67.
4. Wardrop, P., et al., *A temporal bone study of insertion trauma and intracochlear position of cochlear implant electrodes II: comparison of spiral clariontrade mark and HiFocus IItrade mark electrodes banded and Nucleus Contour electrodes.* Hearing Research, 2005. **203**: p. 68-79.
5. Wei, W., K. Xu, and N. Simaan. *A compact Two-armed Slave Manipulator for Minimally Invasive Surgery of the Throat*. in *The first IEEE / RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics (BIOROB'2006)*. 2006. Pisa, Italy.
6. Ikuta, K., K. Yamamoto, and K. Sasaki. *Development of Remote Microsurgery Robot and New Surgical Procedure for Deep and Narrow Space*. in *IEEE International Conference on Robotics and Automation*. 2003.
7. Walker, I. *Some Issues in Creating "Invertebrate" Robots*. in *In the Proceedings of the International Symposium on Adaptive Motion of Animals and Machines*. 2000. Montreal, Canada.
8. Gravagne, I. and I. Walker. *On the Kinematics of Remotely-Actuated Continuum Robots*. in *IEEE International Conference on Robotics and Automation*. 2000.
9. Chirikjian, G.S. and J.W. Burdick, *A Modal Approach to Hyper-Redundant Manipulator Kinematics.* IEEE Transactions on Robotics and Automation, 1994. **10**(3): p. 343-354.
10. Burdick, J.W. and G. Chirikjian, *The Kinematics of Hyper-Redundant Robots*. The IMA Volumes in Mathematics and its Applications, ed. J. Baillieul, S. Sastry, and H. Sussmann. Vol. 104. 1998: Springer-Verlag.
11. Mochiyama, H. and H. Kobayashi. *The Shape Jacobian of a Manipulator with Hyper Degrees of Freedom*. in *IEEE International Conference on Robotics and Automation*. 1999.
12. Mochiyama, H., E. Shimemura, and H. Kobayashi. *Shape Correspondence between a Spatial Curve and a Manipulator with Hyper Degrees pf Freedom*. in *IEEE/RSJ International conference on Intelligent Robots and Systems (IROS`)*. 1998.
13. Angeles, J. and C. Lopez-Cajun, *Optimization of cam mechanisms*. 1991: Kluwer Academic Publishers.
14. Lancaster, P. and L. Rodman, *Algebraic Riccati Equations*. 1995: Oxford Science Publications.
15. Cohen, L., et al., *Improved and Simplified Methods for Specifying Positions of the Electrode bands of a Cochlear Implant Array.* The American Journal of Otology, 1996. **17**: p. 859-865.
16. Ketten, D.R., et al., *In vivo measures of cochlear length and insertion depth of Nucleus cochlear implant electrode arrays.* Ann. Otol. Rhinol. Laryngol., 1998. **107**(12): p. 1-16.
17. Yoo, S.K., et al., *Three-Dimensional Modeling and Visualization of the Cochlea on the Internet.* IEEE Transactions on Information Technology in Biomedicine, 2000. **4**(2): p. 144-151.
18. Juvinall, R. and K. Marshek, *Fundamentals of Machine Component Design*. 3rd ed. 2003: John Wiley & Songs.

Robot-Assisted Prostate Brachytherapy

Yan Yu[1,5], Tarun Podder[1], Yongde Zhang[1], Wan-Sing Ng[6], Vladimir Misic[1], Jason Sherman[1], Luke Fu[1], Dave Fuller[1], Edward Messing[2,4], Deborah Rubens[3,4,5], John Strang[3,4], and Ralph Brasacchio[1]

Departments of
[1] Radiation Oncology
[2] Urology
[3] Radiology and
[4] Surgery,
University of Rochester, Rochester, NY 14642, USA
[5] Department of Biomedical Engineering,
University of Rochester, Rochester, NY 14642, USA
[6] School of Mechanical and Aerospace Engineering,
Nanyang Technological University, Singapore 639798

Abstract. In contemporary brachytherapy procedures, needle placement at the desired target is challenging due to a variety of reasons. A robot-assisted brachytherapy system can improve the needle placement and seed delivery resulting in enhanced patient care. In this paper we present a 16 DOF (degrees-of-freedom) robotic system (9DOF positioning module and 7DOF surgery module) developed and fabricated for prostate brachytherapy. Techniques to reduce needle deflection and target movement have been incorporated after verifying with extensive experiments. Provisions for needle motion and force feedback have been included into the system for improving the robot control and seed delivery. Preliminary experimental results reveal that the prototype system is quite accurate (sub-millimeter) in placing brachytherapy needles.

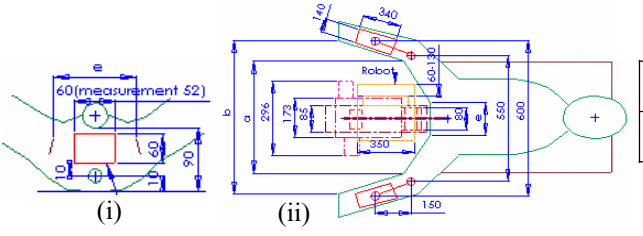
1 Introduction

In traditional brachytherapy procedures, the needles are inserted transperineally under the guidance of transrectal ultrasound (TRUS) images. Both the needle and the TRUS are operated manually; the seeds are deposited using a manual applicator. The needles are inserted through fixed holes of a physical template. Therefore, flexibility and maneuverability of needle insertion are severely limited. Sometimes it is difficult to avoid pubic arch (especially for patients with larger prostates) because the needles can only be inserted straight through the template's holes. The consistency and efficiency of the treatment procedure are highly dependent on the clinicians. In order to assist clinicians, it is important to develop a motorized semi-automated robotic system for prostate brachytherapy. Robotic system will not only accurately place the needle and deliver seeds but also can assist less experienced or skillful surgeons to perform the procedures with higher consistency and efficiency.

Several researchers have been developing robotic systems for prostate brachytherapy procedures [1]-[7]. The needle placement robot system developed by

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 41 – 49, 2006.

Fichtinger et al. [1] and Stoianovici et al. [2] comprised a 3 degree-of-freedom (DOF) Cartesian bridge over the patient, a 2DOF remote center of motion (RCM), and a 1DOF needle insertion with a motorized driver using an axially loaded friction transmission. A 7DOF passive arm was employed between the Cartesian stage and the rest two modules (RCM and needle inserter) to position and orient the needle. Although the stages had sufficient encoders, the needle driver lacked precise encoding of the depth of needle insertion which was performed by a friction drive. In this system, seeds can only be deposited manually. Wei et al. [3],[4] and Wan [5] et al. have developed a robotic system for prostate brachytherapy where they have used an industrial robot to position and orient a single hole template through which a needle can be inserted manually. A separate motorized module is used for operating the US probe. The needle placement robot system designed by Kronreif et al. [6], [7] consisted of two offset x-y stages which allow positioning and orienting the needle over the perineum in 4DOF (two translational and two rotational). The needle can be inserted through the needle guide manually by the clinician.



a (mm)	b (mm)	e (mm)
420-500	450-600	100-120

Fig. 1. Workspace for robotic-assisted brachytherapy (i) front view, (ii) top view

The available workspace for the robot is quite limited while the patient is in lithotomy position for transperineal prostate brachytherapy (Fig. 1). Thus most of the industrial robots may loose dexterity (or may loose DOF) working in severely constrained workspace in the operating room (OR). Currently, hardly any robotic system is available for full functionalities that are required during actual brachytherapy in the OR. In this paper, we present a robotic system (called EUCLIDIAN – Endo-Uro Computed Lattice for Intratumoral Delivery, Implantation, Ablation with Nanosensing) for brachytherapy that is more equipped to perform most of the required procedures automatically (or semi-automatically). We present a detailed design and development of EUCLIDIAN, sequence of operations in clinical procedures, and experimental results to evaluate EUCLIDIAN's accuracy and repeatability.

2 System Design and Development

We have designed and fabricated a more complete robotic system for assisting clinicians during prostate brachytherapy procedures. We collected *in-vivo* data measuring the available workspace for a robot during actual brachytherapy procedures (Fig. 1); this information helped us in designing a compact robotic system that can work efficiently in the severely constrained workspace in the OR.

2.1 Objectives

The main objects of this robotic system are to (1) Increase accuracy of needle placement and seed delivery, (2) Increase avoidance of critical structures (urethra, pubic arch, bladder), (3) Update dosimetry after each needle is implanted, (4) Detect tissue heterogeneities and deformation via force sensing and imaging feedback, (5) Reduce tediousness and assist clinicians, (6) Reduce trauma and edema, (7) Reduce radiation exposure, (8) Reduce learning curve, and (9) Reduce OR time.

2.2 Functional Requirements

The functional requirements of the system are: (1) Quick and easy disengagement in case of emergency, (2) Provision for reverting to conventional manual brachytherapy method at any time, (3) Improvement of prostate immobilization techniques, (4) Provision for periodic quality assurance checking, (5) Updating implant plan after implanting the periphery of the prostate, or most of the needles have been placed and seeds implanted, or any time, (6) A method for the clinician to review and approve the motion plan before the first needle placement, (7) Ability to modulate velocity and needle rotation by automatic feedback control, (8) Visual confirmation by the chosen imaging technique of each seed deposition or the needle tip at the resting position, (9) Steering of the needle by automatic feedback control, (10) Force feedback during needle insertion, and (11) Ease of operation and safety for the patient and OR environment.

2.3 Prototype Robotic System

We have designed and fabricated a prototype robotic system for prostate brachytherapy. It comprises of a 9DOF Positioning and a 7DOF Surgery Module.

2.3.1 Positioning Module

3DOF Cart – the cart can move in x-y horizontal plane and rotate about a vertical axis. It consisted of a base having four wheels capable of rotating about two of its own axes (a horizontal axis and a vertical axis) and a floor locking mechanism; while locking the cart, it drops four legs and lifts the whole system off the wheels. This ensures rigid locking on any floor and provides greater rigidity to the system. Above the base, all the electronic and electrical components including an industrial computer are housed in an enclosure. The cart is brought near the patient's bed to a relatively rough position and locked on the floor by pressing a lever, then the positioning platform is suitably adjusted.

6DOF Platform – it connects the surgery module to the cart. The platform has 3DOF translational motions and 3DOF rotational motions. Thus it can position and orient the surgery module at any location in the 3D space so that the ultrasound (US) probe can be positioned and oriented in the patient's rectum easily and comfortably, and at the same time the needling mechanism should be suitably aligned with the patient's perineum. The vertical lift (y-motion) of the platform is motorized for ease of operation against gravitational effect. The translational motion in horizontal plane are manual, however these joints are unlocked using two solenoids, for safety and stability they are by default locked. The 3DOF rotational motions (roll-pitch-yaw) are achieved by using a spherical joint which can be mechanically locked at a desired orientation.

2.3.2 Surgery Module

2DOF Ultrasound Probe Driver – the US probe (Acuson™ 128xP, Mountain View, CA) can be translated and rotated separately by two DC servo motors fitted with encoders and gear boxes. This enables imaging in transverse plane as well as in sagittal plane, providing capability for improving 3D prostate model for dosimetric planning. Working ranges of motions for US probe are 0-185mm and -91^0 to $+91^0$ in translation and rotation, respectively. The clinician can also drive the US probe manually using the knobs; during this mode the motors are automatically disengaged by the electrical clutches (Fig. 2).The template holder at end of the US probe driver is a provision for manual takeover, if required. The prostate stabilization needle guide can orient the needle at any desired angle in both horizontal and vertical plane resulting in improved stabilization of the prostate and thereby enhanced delivery of the seeds.

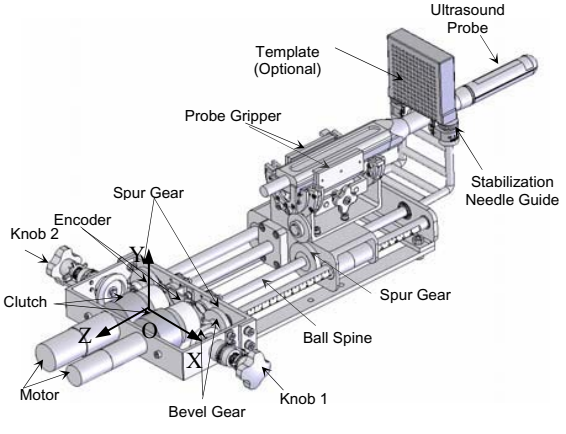


Fig. 2. Ultrasound probe driver

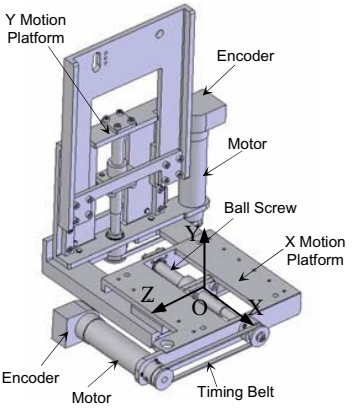


Fig. 3. Gantry robot

3DOF Gantry – it connects the needle driving module to the positioning platform. The gantry has two translational motions (x- & y- direction) and one rotational motion (pitch). The motions are achieved by DC servo motors and optical encoders fitted with the motors (Fig. 3). The working range of motion of the x-direction is 0-62mm and in y-direction is 0-67mm which are sufficient to cover the projected treatment area of a conventional template grid (60mm x 60mm). The rotational range for angulating the needle to avoid pubic arch is -5^0 to $+5^0$. The 3DOF motions of the gantry enable to position and orient the needle at any desired location on the patient's perineum with greater freedom due to the absence of a physical template used in conventional brachytherapy. Once the needle is positioned at the desired location close to the perineum, the needle driver inserts the needle into the patient. The motions of the US probe driver and the rest of the surgery module (gantry and needle driver) are decoupled by making two separate open kinematic chains attached to the same positioning platform.

2DOF Needle Driver – the needle which consists of a hollow cannula and a solid stylet are driven separately by two DC servo motors. The cannula is rotated

continuously or partially using another tiny DC motor. Both the stylet and the cannula are driven from the back (they are pushed); therefore there is no chance of slipping. During actual brachytherapy procedures, the needle traverses through different types of tissues and organs to reach the target in the prostate, and these tissues and organs have different types of boundary conditions that cannot be assessed from experimental data obtained from *ex-vivo* tissue/organ samples. Therefore, *in-vivo* measurement of needle insertion force (Figure 4) [9] is very useful in designing and controlling any robotic system that will work in such a constrained space (Fig. 1). In our design process, we have used this *in-vivo* information. It has been proven that continuous rotation can improve targeting accuracy and can reduce insertion force , whereas partial rotation can increase needle placement accuracy [5],[9].Therefore, we have incorporated the provision for needle rotation in our EUCLIDIAN robot for improving accuracy and seed delivery. To measure and monitor force profiles during the operational procedures, we have installed two single-axis force sensors (Model 13, Honeywell Sensotech, Columbus, OH) each at the proximal ends of the stylet and cannula, and one six-axis force-torque sensor (Nano17, ATI Industrial Auto., Apex, NC) at the distal end of the cannula (Figure 5). Monitoring of these forces is useful in detecting pubic arch interference (PAI) and will help in assessing needle bending. Unlike commercially available seed cartridges (hold maximum 15 seeds), this cartridge can accommodate 35 seeds at a time resulting in less frequent replacement of the cartridge, and thereby reducing potential radiation exposure as well as reducing OR time.

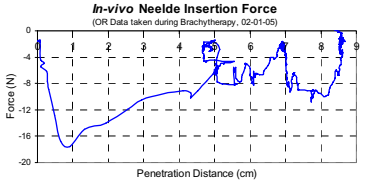


Fig. 4. *In-vivo* needle insertion force measured during actual prostate brachytherapy procedure (conventional) in the operating room

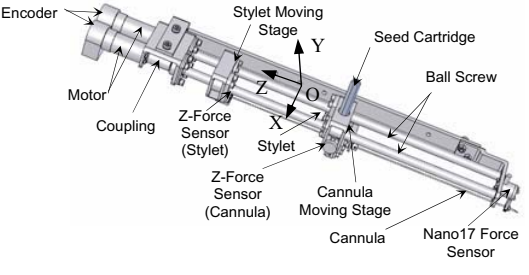


Fig. 5. Needle driver

Seed Pusher – a flat ended style (for beveled-tip brachytherapy needle) is used to push the seed out of the cartridge and to deposit it at the planned locations in the prostate; the stylet's motor is deployed to push the stylet from the proximal end (Fig. 5). All the sequence of motions during the seed delivery is fully automatic; however, at any time the clinician can interrupt the motion. The force sensor at the proximal end of the stylet monitors the force profile on the stylet and thereby confirms the seed removal from the cartridge to the cannula; experiments showed that on an average about 2.5N force is required to push the seed out of the cartridge. This is also checked by monitoring the stylet's motor current.

User's Pendant – the handheld pendant with 10 buttons provides the surgeon the freedom to take over the control of the surgery module at any desired time. From the

user's pendant the surgeon can control the needle insertion, needle rotation, x-y movement of the gantry and system abort.

The EUCLIDIAN system is quite compact to work in the constrained workspace available during prostate brachytherapy procedures (Figs. 6 & 7). The overall maximum dimensions (length x width x height) of the surgery module are 510mm x 290mm x 235mm; width at patient's end (distal end) of US Probe Diver is 100mm and that of Needle driver is 60mm. The gross weight of the surgery module is about 9kg, which can be reduced significantly by using plastic/nylon instead of presently used surgical grade stainless steel and aluminum. All the motors are fitted with optical encoders (MicroMo Electronics, Inc., Faulhaber Group, Clearwater, FL) which provide final motion resolutions (considering gear ratios and screw leads) of 0.0007mm for gantry x-y translations, 0.004mm for stylet and cannula motions; 0.005mm and 0.06^0 for US probe translation and rotation, respectively.

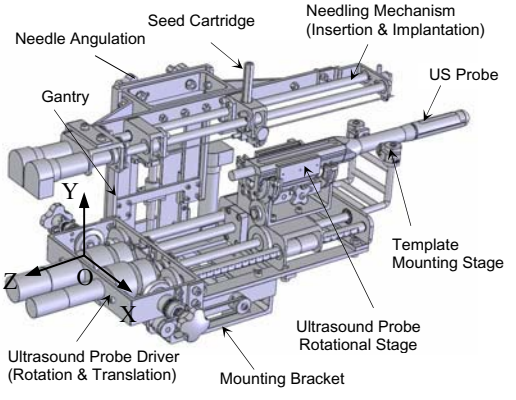


Fig. 6. Assembled surgery module

Fig. 7. Prototype robotic system

3 Clinical Procedures Using the Prototype Robotic System

Clinical work flow of the prostate brachytherapy procedures using the robotic system is segmented into seven steps as described below.

In setup state, the EUCLIDIAN is initialized and patient information is entered into the computer by the user. Then the TRUS (transrectal ultrasound) is moved to scan the prostate in transversal plane and the images at a desired interval are saved. In the next step, modeling state, the TRUS images are used in delineating the prostate boundary, urethra, pubic bone, rectum, and seminal vesicle. Then a 3D model of the prostate is generated automatically.

Next, this 3D model of the prostate is used for dosimetric planning to obtain the desired coordinates of the radioactive seed distribution. The designed software can display the planned iso-dose contours, needle position and seed locations in 3D. This

provides the clinicians a useful visualization of the whole treatment plan and if required, the clinicians can edit the plan.

Once the plan is approved by the clinicians, a single needle is inserted into the patient according to the plan. At this stage the TRUS is employed in sagittal plane to tract the needle location. To ensure patient's safety, this needle insertion is performed in a sequential order, i.e. first, the gantry moves in x-y direction to bring the tip of the needle close to the perineum of the patient; second, the gantry is stopped and then the needling mechanism pushed the needle (stylet & cannula together) into the patient unto a predefined depth; third, the clinician uses the user's pendant to insert the needle up to the final depth. The last stage of user's takeover is to ensure patient's safety as well as to accommodate any change in planned depth that may be required due to tissue/organ deformation or needle deflection. After the needle insertion, the system is in implanting state. Seed is loaded from cartridge and implanted according to the plan, and the needle is withdrawn.

As seeds being delivered, the system can go into a validation state to validate the latest dosimetry, or skip the validation and return to the needling state. In validation state, the prostate is scanned and seeds are located in order to update the dosimetry for validation purpose. If the clinicians find that the dosimetry is not satisfactory, the system can return to the planning state to make adjustment for the remaining seeds to be inserted.

The patient safety is ensured by a variety of hardware (sensors and stops) and software checks. An emergency button on the top of the cart can be used to stop the whole system in case of emergency. The clinician can move the needle and gantry at any desired speed by using the user's pendant. He can also abort the any motion at any time using the pendant. Although the system can execute coordinated (or simultaneous) motion of all the axes, we preferred to have sequential motions of the critical components such as needle insertion, seed delivery, and needle withdrawal for patient's safety reasons. However, to optimize total operating time some of the motions (US probe and gantry) are coordinated. Sterilization of needle passages and seed passages have been ensured by incorporating sterilized inserts. In this prototype system, these inserts are permanent; however, for routine use they can be of

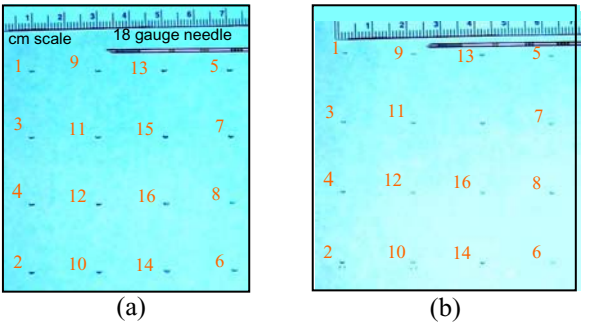


Fig. 8. Position and size of holes (a) after one hundred penetrations at each location by moving the needle from one location to the next location after each penetration, and (a) after single penetration at each location

disposable type. Most of the parts of the EUCLIDIAN robot are isolated from patient's blood and fluids by using leather jacket for the positioning module and drapes for the TRUS driver, gantry, and needling mechanism. The cleaning and decontamination of the relevant portions of the system is performed the same way as that currently practiced in the OR.

4 Results and Discussions

To evaluate the system's accuracy and repeatability we performed some preliminary experiments. The system was run in a pattern (as shown in Fig. 8) for about 2 hours at a time. An 18 gauge beveled-tip brachytherapy needle was inserted into a graph paper pasted on a block of foam. There were a total of 16 penetration locations in a 60mm x 60mm area. The gantry moved the needle driver in a sequence (1-2-3- …-16), as shown in Fig. 8(a)) so that the mechanical systems go through back and forth motion to experience backlash/slack errors in the motion transfer trains/linkages. After a single penetration at each location the needle was moved to the next location, and this was repeated for 100 times. The x and y speed of the gantry was 10mm/s and the needle speed was 50mm/s. To assess the error and deviation we have also done the same for a single insertion at each location (Fig. 8(b)). It was very difficult to distinguish the deviation of the penetration holes even after 100 runs. We magnified the image of the holes by 5 times to measure the errors. The average errors in x and y motions that the system (needle tip) encountered after 100 insertions was ±0.15mm with repeatability of ±0.05mm.

5 Remarks and Future Work

This paper presented the design and development of a compact robotic system for prostate brachytherapy treatment with radioactive permanent seed deposition. Numerous techniques and sensors were incorporated into the prototype to improve the needle insertion accuracy, and thereby reducing the seed delivery errors; this ensures improved radiation dose delivery to the patient. The system can be operated in both automated and semi-automated modes. In case of any emergency the system can be stopped instantaneously and the clinicians can switch to conventional manual mode. The patient safety is ensured by a variety of hardware (sensors and stops) and software checks. Accuracy and repeatability of the prototype system were tested and they were found to be very accurate. We are performing more rigorous experiments with phantoms to evaluate the overall performance of the system. In addition to US imaging modality we are considering the use of CT (computed tomography) as more precise imaging modality. Once the phantom studies are completed we intend to start clinical trials.

Acknowledgements

This work is supported by the National Cancer Institute, under grant R01 CA091763.

References

1. Fichtinger, G., DeWeese, T. L., Patriciu, A., et al. System for Robotically Assisted Prostate Biopsy and Therapy with Intraoperative CT Guidance. Acad. Radio., Vol. 9, (2002) 60-74.
2. Stoianovici, D., Cleary, K., Patriciu, A., et al. AcuBot: A Robot for Radiological Percutaneous Interventions. IEEE Trans. on Rob. and Auto., Vol. 19(5), (2003) 927-930.
3. Wei, Z., Wan, G., Gardi, L., et al. Robot-assisted 3D-TRUS Guided Prostate Brachytherapy: System Integration and Validation. Medical Physics, Vol. 31(3), (2004) 539-548.
4. Wei, Z., Wan, G., Gardi, L., Downey, D.B., and Fenster, A.: Robotic Aided 3D TRUS Guided Intraoperative Prostate Brachytherapy. SPIE, Vol. 5367, (2004) 361-370.
5. Wan, G., Wei, Z., Wan, G., Gardi, L., Downey, D.B., and Fenster, A.: Brachytherapy Needle Deflection Evaluation and Correction. Medical Physics, Vol. 32(4), (2005) 902-909.
6. Kettenbach, J., Kronreif, G., Figl, M., et al. Robot-assisted Biopsy using Computed Tomography-guidance: Initial Results from *In Vitro* Tests. Investigative Radiology, Vol. 40, (2005) 219-228.
7. Kettenbach, J., Kronreif, G., Figl, M., et al. Robot-assisted Biopsy using Ultrasound Guidance: Initial Results from *In Vitro* Tests. European Radiology, Vol. 15, (2005) 765-71.
8. Podder, T.K., Clark, D.P., Sherman, J., Messing, E.M., Fuller, D., Rubens, D.J., et al. In vivo Motion and Force Measurement of Surgical Needle Intervention during Prostate Brachytherapy. Medical Physics Journal, Vol. 33(8), (2006).
9. Podder, T.K., Liao, L., Sherman, J., Misic, V., Zhang, et al. Assessment of Prostate Brachytherapy and Breast Biopsy Needle Insertions and Methods to Improve Targeting Accuracy. The IFMBE Proc. of the Int. Conf. on Biomed. Eng. (ICBME), Vol. 12, (2005).

Design and Validation of an Image-Guided Robot for Small Animal Research

Peter Kazanzides[1], Jenghwa Chang[3], Iulian Iordachita[1], Jack Li[2], C. Clifton Ling[3], and Gabor Fichtinger[1]

[1] Department of Computer Science, Johns Hopkins University, USA
pkaz@jhu.edu
[2] Department of Mechanical Engineering, Johns Hopkins University, USA
[3] Medical Physics Department, Memorial Sloan Kettering Cancer Center, USA*

Abstract. We developed an image-guided robot system to achieve highly accurate placement of thin needles and probes into in-vivo rodent tumor tissue in a predefined pattern that is specified on a preoperative image. This system can be used for many experimental procedures where the goal is to correlate a set of physical measurements with a corresponding set of image intensities or, more generally, to perform a physical action at a set of anatomic points identified on a preoperative image. This paper focuses on the design and validation of the robot system, where the first application is to insert oxygen measurement probes in a three-dimensional (3D) grid pattern defined with respect to a PET scan of a tumor. The design is compatible with CT and MRI, which we plan to use to identify targets for biopsy and for the injection of adenoviral sequences for gene therapy. The validation is performed using a phantom and includes a new method for estimating the Fiducial Localization Error (FLE) based on the measured Fiducial Distance Error (FDE).

1 Introduction

We are investigating non-invasive methods for identifying hypoxic (oxygen deficient) cells in tumors. This is important because hypoxic cells are resistant to radiation treatment and therefore treatment can be improved by tailoring the radiation dosage directed at them. By measuring the tissue oxygen tension (pO_2) level of the cells using an Oxylite probe[1] and correlating these measurements with Positron Emission Tomography (PET) scan data, it is possible to verify the efficacy of PET scans for locating hypoxic cancer cells prior to radiation treatment[2]. Initially, we used a manual method to verify the correlation between PET scan data and pO_2 measurements for tumors on rodents. The procedure was to place the anesthetized rodent inside a foam-filled bed that contained a template with PET-compatible markers. After the PET scans, the entire rodent and bed assembly were placed beneath a passive fixture that held

* We thank Emese Balogh, Anton Deguet, Russell Taylor (JHU), Bixiu Wen, Pat Zanzonico, Andrei Pugachev, Qing Chen, Jose Campa (MSKCC) for their assistance. This work is supported by NIH RO1 CA84596 and NSF EEC 9731748.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 50–57, 2006.

the Oxylite probe. A set of measurement targets (a "track") on the PET image was selected and the corresponding template hole was located. After the skin and tumor were punctured with a needle, the Oxylite probe was manually advanced through the template hole to measure the pO_2 level at each point along the track. This procedure was repeated for multiple template holes (tracks). Our goal was to design and build a robot system to automate this time-consuming and labor-intensive procedure, while eliminating the constraint that measurement tracks be defined only through template holes. The desired accuracy of the robot system was specified to be 0.25 mm or better, not including the error due to the imaging system.

2 System Design

2.1 Hardware Design

The robot system consists of a mobile cart that houses the electronics and provides a table top for the four axis robot and display (see Fig. 1). The robot is composed of a two degree-of-freedom X-Y horizontal platform and two vertical slides (Z1, Z2). A horizontal arm is mounted on the first vertical slide (Z1) and provides an attachment for either a registration probe or cannula. The second vertical slide (Z2) is attached to the first vertical slide and contains a probe holder. This allows the system to insert the Oxylite probe through the cannula and into the tumor. Note that in this case, the Z1 axis positions the cannula near the skin surface and the Z2 axis drives the measurement probe to the target.

The rodent bed (Fig. 2) fits inside the limited bore of the small animal imaging scanners and mounts onto the robot X-Y platform. It includes markers for the registration between image and robot coordinates. The markers are mounted on an adjustable bridge so that they can be positioned over the target region (tumor) and within the scanner field of view. The bridge is removed after registration to enable access to the rodent. As shown in Fig. 2, we initially used the Acustar® marker system[3], donated by Z-Kat Inc. (Hollywood, Florida, USA), for the CT, MRI and robot markers and a separate set of support tubes (offset by a known amount) for the radioactive PET markers. We are currently using a simpler marker system that consists of four small hemispheres (3 mm diameter), drilled into the adjustable bridge. These hemispheres are filled with an appropriate contrast agent prior to imaging. After imaging, the contrast agent is removed and the holes are physically located by the robot. For this procedure, the cannula is replaced by a registration probe, which is guided to the markers using a force control mode[4]. Force control is possible because the system contains a two-axis sensor (XY) beneath the rodent bed and a single-axis sensor (Z1) near the attachment mechanism for the registration probe and cannula.

The robot controller consists of a rackmount computer connected via Ethernet to a DMC-2143 controller board and AMP-20540 power amplifier (Galil Motion Control, Rocklin, CA, USA). The controller provides low-level servo control of the four robot motors. Application software on the PC sends position goals via Ethernet to the controller, which then moves each joint to its goal. The power

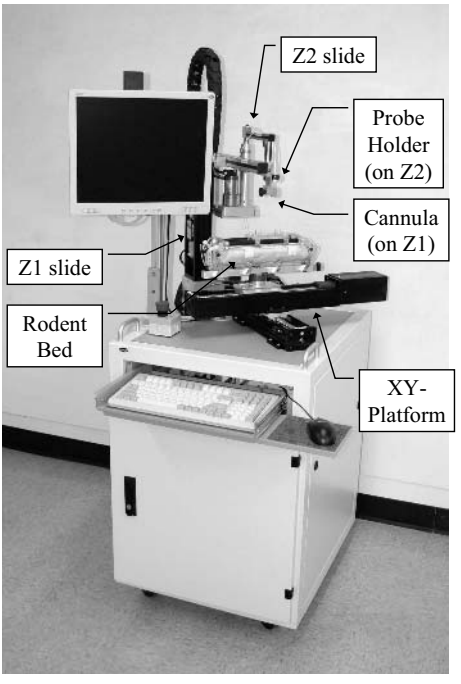


Fig. 1. Robot system

Fig. 2. Rodent bed with marker bridge

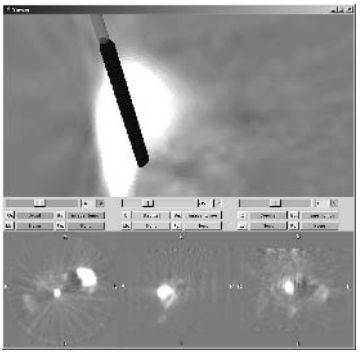


Fig. 3. Measurement track in PET

amplifier provides eight analog inputs, three of which are used for the interface to the force sensors. For the XY force sensor, we used the DX-480 from Bokam Engineering (Santa Ana, CA, USA) with a custom amplifier board that fits inside a shielded metal case below the sensor body. The Z1 force sensor is an L2357 sensor manufactured by Futek, Inc. (Irvine, CA, USA).

2.2 Software Design

We developed the application software, using 3D Slicer (www.slicer.org), to guide the user through most steps of the procedure, which are:

1. Place anesthetized tumor-bearing rodent in rodent bed.
2. Place rodent bed in scanner and obtain image data.
3. Move rodent bed to robot system and load image data into computer.
4. Register image data to robot:
   (a) Manually guide robot's registration probe into contact with each of the four markers.
   (b) Use a semi-automatic image processing procedure to locate the corresponding image markers.
   (c) Perform an optimization to determine the transformation that aligns the two sets of marker positions (robot and image).

5. Remove registration probe from Z1 axis and attach cannula.
6. Attach measurement (Oxylite) probe to Z2 axis and zero its position.
7. Identify target regions (sets of vertical tracks) in the image (see Fig. 3).
8. Transform the track points to robot coordinates and move the robot so that it positions the cannula at the entry point of the first probe track.
9. Prompt the user to manually puncture the skin by inserting a needle through the cannula and then removing it.
10. Command the robot to move the measurement probe through the cannula and into the tumor. The robot moves the probe vertically inside the tumor in user-defined increments (typically 1 mm), recording data at each position.
11. When the measurement probe reaches the end of the current track, command the robot to retract the probe back inside the cannula, move it to the starting point of the next track, and repeat the above cannula insertion and probe measurement sequence until the entire grid pattern has been traversed.

The manual guidance feature is achieved by a force control algorithm that uses nonlinear gains to provide fine positioning without sacrificing maximum motion speed[4]. A simplified representation of the nonlinear force control equation is given by:

$$V_i = F_i * G_i * (1 - \cos(9 * F_i))$$

where V_i, F_i and G_i are the commanded velocity, measured force and maximum admittance gain for axis i, respectively. The multiplication by 9 ensures that the maximum force reading (10 Volts) produces the largest commanded velocity because cos(90) is 0. The nonlinear factor $1 - \cos(9 * F_i)$ is relatively flat for low forces but has a higher slope for large forces. The actual implementation includes a deadband (so that sensor noise does not cause unwanted motion) and another nonlinear factor that is a function of the distance to the travel limit (so that the robot slows down as it approaches the limit).

The transformation ${}^R_I T$ between image coordinates and robot coordinates is obtained by performing a best-fit estimation between the two 3D point sets (4 image markers and 4 robot markers). There are several closed-form solutions that minimize a cost function that is the sum of the squared L2-norms, where ${}^R P_k$ and ${}^I P_k$, $k = 1 \ldots N$, are the robot and image marker positions, respectively ($N = 4$ in our case):

$$Cost = \sum_{k=1}^{N} \left|\left| {}^R P_k - {}^R_I T \; {}^I P_k \right|\right|^2 \tag{1}$$

Due to a somewhat arbitrary implementation decision, we instead compute an initial transformation using three of the markers and then iteratively estimate the final transformation using a variation of Powell's Method [5]. This minimizes a cost function as given in equation (1), except that the L2-norms are not squared. Although our method has the potential advantage of being less sensitive to outliers (because the errors are not squared), it has the disadvantage of being incompatible with much of the literature, which assumes the formulation

of equation (1). For example, the Fiducial Registration Error (FRE) is defined based on the minimum value of equation (1):

$$FRE = \sqrt{\frac{1}{N}\sum_{k=1}^{N} \left|\left|{}^{R}P_k - {}^{R}_{I}T_i \; {}^{I}P_k\right|\right|^2} \tag{2}$$

To avoid confusion, the results of the phantom experiments reported in the next section were computed (offline) using the standard technique described by Arun[6] and modified by Umeyama[7], instead of the iterative technique. In our testing, we noted minor differences between the registrations obtained with these two methods and we intend to replace the iterative technique with this closed-form technique.

3 Phantom Experiments

We performed several tests using a phantom to evaluate the accuracy of the overall system, as well as the accuracy of the major subsystem components (imaging system and robot). Following the terminology of [3], we focus on the following categories of error:

Fiducial localization error (FLE): error in determining the position of a marker in image coordinates (FLE-I) or robot coordinates (FLE-R).

Target registration error (TRE): mean error for locating markers that were not used for registration. This is most relevant to the application.

3.1 Phantom Design

We designed a phantom (Fig. 4) that has 20 small hemispherical holes (1-20), 2 cylindrical holes (C1-C2) and 4 large registration holes (R1-R4) arranged at 5 different heights. The 2 cylindrical holes and 4 large registration holes (R1-R4) were not used for any of the tests reported here and will not be discussed further. The 20 small holes have a diameter of 3 mm and are therefore equivalent to the markers on the rodent bed. Four of these holes (1, 3, 10, 11) are arranged in the same pattern as the four markers on the rodent bed. We chose Delrin because it is compatible with all image modalities of interest (PET, CT, MRI). The phantom is 120mm x 50mm x 45mm, which is small enough to fit inside all small animal scanners. It was machined on a CNC machine with a known accuracy of ± 0.0005 inches (± 0.0127mm). Considering the material properties of Delrin, we estimate the overall accuracy of the phantom to be ± 0.05mm. Because our "hole finding" procedure with the robot is a manual task involving hand-eye coordination, we darkened the edges of the holes to obtain sufficient visual contrast.

For microPET imaging of the phantom, the hemispherical holes were filled with contrast agent (radioactive tracer). After scanning, the application software was used to find the centroid of each marker in the image. For the robot measurements, the robot's registration probe was manually guided to each accessible hole. The data was analyzed to determine the Fiducial Localization Errors (FLE-I and FLE-R) and the Target Registration Error (TRE).

3.2 Fiducial Localization Error (FLE)

FLE can be simply defined as "the error of determining the positions of the markers" [3], but it can be difficult to measure directly. One approach is to infer FLE from FRE, using the following approximation (see [8]):

$$FLE = \sqrt{\left(\frac{N}{N-2}\right) FRE^2} \tag{3}$$

This is especially useful with an accurately-machined (CNC) phantom because it can be individually applied to each measurement subsystem (image and robot). Furthermore, the phantom can contain a large number of markers and thereby produce a robust estimate of FLE.

We developed an alternate method for estimating FLE based on the Fiducial Distance Error (FDE) which we define as the difference, for each pair of markers, between the measured distance and the known distance. For example, if P_{ab} is the measured distance between markers a and b (in image or robot coordinates) and C_{ab} is the distance between those markers in CNC coordinates, the FDE is $|P_{ab} - C_{ab}|$. Note that for N markers, there are a total of $N(N-1)/2$ measurements. The relationship between FLE (assumed identical at each marker) and the average FDE depends on the geometrical arrangement of the markers. We performed simulations to obtain an empirical relationship. Each simulated data set was created by adding zero-mean Gaussian noise, with a specified standard deviation (applied isotropically), to the CNC coordinates of each phantom marker. For each standard deviation, we created 10,000 simulated data sets and computed the average FDE. We performed separate simulations for the image markers and the robot markers because even though we used the same phantom, 14 markers were in the PET field of view and 16 markers were accessible by the robot. In both cases, we obtained a linear relationship between FLE and FDE, with the ratio FLE/FDE approximately equal to 1.5.

3.3 Target Registration Error (TRE)

It is first necessary to find the four registration markers in PET and robot coordinates and compute the transformation between these two coordinate systems. The TRE is then determined by transforming all other markers (not used for registration) into a common coordinate system and computing the distance between each set of corresponding points; for example, if P_5 and R_5 are the positions of point 5 in PET and robot coordinates, respectively, and ${}^R_I T$ is the transformation from PET (image) to robot coordinates, the TRE for point 5 is given by $\left|R_5 - {}^R_I T P_5\right|$.

TRE is the most meaningful measurement of the overall system accuracy because it estimates the accuracy with which the robot can position the instrument (Oxylite probe) at the location identified on the PET image. It is important, however, to understand the limitations of the test conditions. The phantom tests do

not include errors due to instrument compliance/bending or to the motion of the target (tumor). In the case of the fiber-optic Oxylite probe, the error due to bending can be significant. Furthermore, the TRE value does not include any error due to the tool drive axis (Z2).

4 Results

The hemispherical markers in the phantom were filled with 5 μl of ^{18}F-FDG (26.5 μCi). Only 14 markers (1-14) were visible in the microPET scanner's limited field of view. On the robot, it proved to be too difficult to see the 4 deepest markers (2, 7, 13, 17) and therefore they were eliminated from the testing.

4.1 PET Image Fiducial Localization Error (FLE-I)

One microPET scan of the phantom was acquired and the application software was used to find all 14 imaging markers (filled holes) in the field of view. We used a corrected value (1.229 mm) for the PET slice spacing that had been experimentally determined during earlier testing.

The 14 markers resulted in 91 computed distance errors and yielded an FDE-I of 0.172 $\pm$0.121 mm (mean $\pm$ standard deviation). The largest distance error was 0.497 mm. Our simulations produced the empirical relationship FLE-I = 1.49*FDE-I, which estimates FLE-I to be 0.26 mm.

For comparison, we applied the standard least-squares registration technique [6][7] to all 14 imaging markers. The resulting FRE was 0.236 mm, which according to equation (3), with $N = 14$, estimates FLE-I to be 0.26 mm.

4.2 Robot Fiducial Localization Error (FLE-R)

The robot was used to locate the 16 accessible markers in three different trials. Each trial yielded 120 computed distance errors, with an FDE-R value (over all three trials) of 0.121 $\pm$ 0.096 mm (mean $\pm$ standard deviation). The largest distance error was 0.455 mm. Our simulations produced the empirical relationship FLE-R = 1.51*FDE-R, which estimates FLE-R to be 0.18 mm.

Once again, we compared our result to the one obtained by applying equation (3) to the FRE from the least-squares registration of all 16 markers. For the data obtained from the three trials, the FRE was 0.168, 0.127, and 0.117 mm, for a mean value of 0.137 mm. This estimates FLE-R to be 0.15 mm.

4.3 Target Registration Error (TRE)

There were 11 markers, including all 4 reference markers, that were found in the PET image and by the robot. Therefore, it was possible to register the PET data set to each of the 3 robot data sets and compute TRE for the 7 target markers (see Table 1).

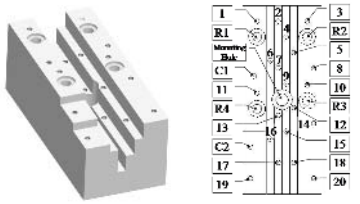


Fig. 4. Phantom

Table 1. Target Registration Error, mm

Trial	FRE	TRE		
		Mean	Std Dev	Max
1	0.217	0.305	0.098	0.447
2	0.200	0.293	0.088	0.395
3	0.227	0.279	0.105	0.475
Mean	0.215	0.292	0.097	

5 Conclusions

We completed the design of an image-guided robot system to assist with cancer research and performed phantom experiments to measure its accuracy. Our results indicate a mean positioning accuracy (TRE) of 0.29 mm, with the accuracy of the robot system alone well within the 0.25 mm specification. The results do not include errors due to the instrument drive unit (Z2 axis), instrument compliance/bending, or motion of the target (tumor). The largest error source is marker localization in the PET image (FLE-I), followed by marker localization by the robot (FLE-R). We introduced a new method for estimating FLE based on distance errors (FDE) and found good agreement with the existing method based on FRE. Our results indicate that this robot system should improve the efficiency and accuracy of needle-based procedures for in-vivo measurement, biopsy, and injection in small animals.

References

1. Urano, M., Chen, Y., Humm, J., Koutcher, J., Zanzonico, P., Ling, C.: Measurements of tumor tissue oxygen tension using a time-resolved luminescence-based optical oxylite probe: Comparison with a paired survival assay. Radiation Research **158** (2002) 167–173
2. Cherry, S., Shao, Y., et al.: MicroPET: A high resolution PET scanner for imaging small animals. IEEE Trans. on Nuclear Science **44** (1997) 1161–1166
3. Maurer, C., Fitzpatrick, J., Wang, M., Galloway, R., Maciunas, R., Allen, G.: Registration of head volume images using implantable fiducial markers. IEEE Trans. on Medical Imaging **16** (1997) 447–462
4. Kazanzides, P., Zuhars, J., Mittelstadt, B., Taylor, R.: Force sensing and control for a surgical robot. In: IEEE Intl. Conf. on Robotics and Automation, Nice, France (1992) 612–617
5. Press, W., Teukolsky, S., Vetterling, W., Flannery, B.: Numerical Recipes in C. 2 edn. Volume 1. Cambridge University Press, Cambridge, UK (1992)
6. Arun, K., Huang, T., Blostein, S.: Least-squares fitting of two 3-D point sets. IEEE Trans. on Pattern Analysis and Machine Intelligence **9** (1987) 698–700
7. Umeyama, S.: Least-squares estimation of transformation parameters between two point patterns. IEEE Trans. Pattern Anal. and Mach. Intell. **13** (1991) 376–380
8. Fitzpatrick, J., West, J., Maurer, C.: Predicting error in rigid-body point-based registration. IEEE Trans. on Medical Imaging **17** (1998) 694–702

GPU Based Real-Time Instrument Tracking with Three Dimensional Ultrasound

Paul M. Novotny[1], Jeffrey A. Stoll[2], Nikolay V. Vasilyev[3], Pedro J. del Nido[3], Pierre E. Dupont[2], and Robert D. Howe[1]

[1] Division of Engineering and Applied Sciences, Harvard University, MA USA
[2] Department of Aerospace and Mechanical Engineering, Boston University, MA USA
[3] Department of Cardiovascular Surgery, Children's Hospital Boston, MA USA

Abstract. Real-time 3D ultrasound can enable new image-guided surgical procedures, but high data rates prohibit the use of traditional tracking techniques. We present a new method based on the modified Radon transform that identifies the axis of instrument shafts as bright patterns in planar projections. Instrument rotation and tip location are then determined using fiducial markers. These techniques are amenable to rapid execution on the current generation of personal computer graphics processor units (GPU). Our GPU implementation detected a surgical instrument in 31 ms, sufficient for real-time tracking at the 26 volumes per second rate of the ultrasound machine. A water tank experiment found instrument tip position errors of less than 0.2 mm, and an *in vivo* study tracked an instrument inside a beating porcine heart. The tracking results showed good correspondence to the actual movements of the instrument.

1 Introduction

Real-time three-dimensional ultrasound (3DUS) is a viable tool for guiding surgical procedures [1]. This visualization technique may enable a range of new minimally invasive procedures in cardiac surgery. For example, beating heart intracardiac surgery is now possible with the use of 3DUS and minimally invasive instruments [2][3]: ultrasound permits visualization through the opaque blood pool in the heart, and the advent of real-time 3DUS overcomes many of the difficulties with 3D spatial perception in conventional 2D ultrasound [1]. These procedures eliminate the need for a cardiopulminary bypass and its well documented adverse effects, including delay of neural development in children, mechanical damage from inserting tubing into the major vessels, increased stroke risk, and significant decline in acute and chronic cognitive performance [4][5][6].

Initial animal trials highlighted several obstacles to clinical implementation of ultrasound-guided intracardiac surgery [2][3]. One such limitation is instrument perception, due to the distorted appearance of stiff materials under 3DUS, including high noise levels, shadowing, and a variety of artifacts. To address this issue, researchers have developed techniques to detect instruments in ultrasound.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 58–65, 2006.

By detecting instruments as they move within the ultrasound image it is possible to highlight the position of the instrument for the surgeons. Eventually, this will allow surgeons to more accurately control the instruments as they perform a surgical task. For example, Vitrani *et al.* [7][8] tracked the fingers of a endoscopic grasper with 2D ultrasound. In addition, there have been numerous reports of tracking surgical needles with ultrasound [9]. These techniques, however, use 2D ultrasound or track instruments in homogeneous static tissue, which limits the applicability to the heterogeneous dynamic environment found in intracardiac procedures. Stoll *et al.* showed that instruments could be tracked in 3DUS using passive markers[10]. This implementation employed singular value decomposition for instrument shaft detection that was unable to distinguish instruments from tissue or other objects. It was also incapable of tracking instruments faster than 1.5Hz and therefore unsuitable for beating heart procedures.

In this work we present a technique capable of detecting instruments used in minimally invasive procedures. We build upon previous work by combining two methods: Radon transform-based instrument shaft detection in 3DUS [11] and passive markers for instrument tip localization [12]. These two methods are complimentary, and we describe a hybrid approach for complete 6 degree of freedom instrument tracking. In addition, we demonstrate that this hybrid technique is well-suited for implementation on powerful personal computer graphics processor units (GPUs). By using a GPU based approach, the complete system is capable of real-time tracking at the rate of the 26 volumes per second produced by the 3DUS machine. The tracking method is validated in water tank and *in vivo* experiments.

2 Methods and Materials

To attain real-time tracking of instruments within an 3DUS image, we utilized techniques that were especially amenable to high performance implementations. This consideration is especially important due to the high data rates of the real-time 3DUS, where volumes (204x48x148 voxels) are produced at a rate of 25-28 Hz. This high data rate requires that real-time algorithms must handle 40 MB of data every second. For this challenge, we build upon work presented by Novotny *et al.* [11], that introduces a high performance instrument shaft detection algorithm. We also incorporate passive markers similar to those presented in [12], but pay careful attention to marker design for increased performance.

2.1 Instrument Tracking

The first step is to detect the axis of the instrument. Generally, minimally invasive instruments such as those used in intracardiac procedures are fundamentally cylindrical in shape (Fig. 1). Graspers, scissors, needle drivers, and clip deployment devices are all attached to a long cylindrical tube. A modified form of the Radon transform has been shown to be capable of identifying these long straight cylinders (Fig. 2), a feature that is not found in cardiac ultrasound images [11].

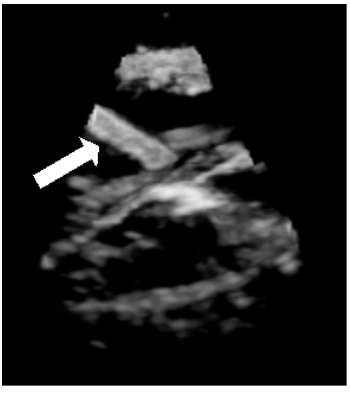


Fig. 1. *In vivo* 3D ultrasound image of a porcine heart with a 5 mm instrument inserted into the right atrium

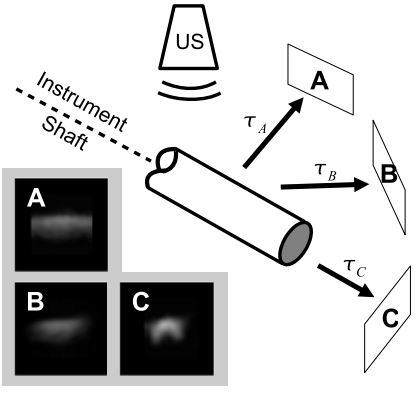


Fig. 2. Example of the modified Radon transform. Each image (A-C) is a projection of the ultrasound image along the direction (τ_A-τ_C).

The algorithm identifies maximums of the the equation:

$$\breve{g}(\theta, \phi, u, v) = \int g(s\tau(\theta, \phi) + u\alpha(\theta, \phi) + v\beta(\theta, \phi))\, ds \tag{1}$$

where $\breve{g}$ is referred to as the modified Radon space of the ultrasound volume, g. Each point in $\breve{g}$ corresponds to the integral along a three dimensional line defined by the four parameters θ, ϕ, u and v. The angular parameters, θ and ϕ, define an orthonormal basis composed of α, β, and τ that are defined as

$$\tau = \begin{pmatrix} cos\theta cos\phi \\ sin\theta cos\phi \\ sin\phi \end{pmatrix}, \ \alpha = \begin{pmatrix} -sin\theta \\ cos\theta \\ 0 \end{pmatrix}, \ \beta = \begin{pmatrix} -cos\theta sin\phi \\ -sin\theta sin\phi \\ cos\theta \end{pmatrix}. \tag{2}$$

Two positional parameters, u and v, are also used to fully define a 3D line (Fig. 3). Identifying lines in a 3D volume now becomes a problem of finding local maximums of $\breve{g}$. By finding the maximum value of $\breve{g}$, the axis of the instrument in 3D space is implicitly defined by the parameters (θ, ϕ, u, v).

This algorithm is especially well suited for implementation on parallel architectures such as modern graphics processing units. In its original implementation it detected instruments in an ultrasound volume in 0.6 s [11]. To improve upon this previous implementation, we both refined the search algorithm that identifies maximums of Eqn. 1 and also used the information from previous volumes to seed the tracking of the current volume.

To this end, we start by calculating Eqn. 1 for evenly spaced points throughout the volume space g. Spatially, the volume is sampled at 5 voxel increments in x, y, and z. For angles θ and ϕ, Eqn. 1 is sampled in 10 degree increments. Due to symmetry, the angles are only sampled from 0 to 180 degrees. This search constitutes the initialization of the instrument tracking, as the entire volume

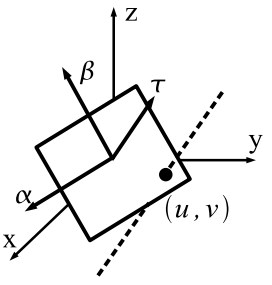


Fig. 3. Schematic of the modified Radon transform in 3D. The transform integrates along a line (shown as dashed), defined by α, β, and τ.

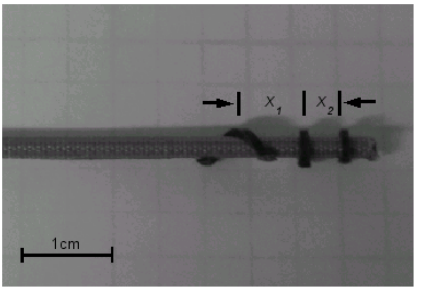


Fig. 4. Picture of an instrument with passive markers. The instrument tip and roll angle is calculated using the distances x_1 and x_2.

is searched for the instrument. For 148x48x208 voxel volumes, this results in 408,000 iterations of Eqn. 1.

For subsequent frames, the tracking algorithm confines its search space to an area centered on the location found in the previous frame. Since the ultrasound volumes are updated at 25-28 Hz, this search space can be fairly small. In our trials, we empirically found that limiting the search space to ± 5 voxels spatially in the x, y, and z directions and ± 10 degrees around the angles θ and ϕ found in the previous frame was sufficient to capture typical surgical movements.

To enable real-time tracking, the instrument axis tracking algorithm is calculated on the GPU where the 3DUS volume is stored in texture memory. Each calculation of Eqn. 1 is performed by a pixel shader on the GPU. These pixel shaders take advantage of high performance volume interpolation and vector manipulation built into the GPU. In addition, the GPU used in the implementation (7800GT, nVidia Corp, Santa Clara, CA) has 16 pipelines that calculate Eqn. 1 simultaneously for 16 different positions in the $\breve{g}$ space. Once the $\breve{g}$ space has been sufficiently sampled, the maximum identifies the position of the instrument axis.

2.2 Passive Markers

Once the axis of the instruments is found, it is necessary to detect the final two degrees of freedom of the instrument (tip position and roll angle) to fully define its position and orientation. To this end, we build on work first introduced by Stoll *et al.* [12]. Here we use a new marker design, shown in Fig. 4. To produce distinct elements, 800 μm polyurethane foam was wrapped around the instrument shaft as shown in Fig. 4. Uncoated metals such as the stainless steel used for surgical instruments are highly reflective in ultrasound. To ensure that the instrument is visible in ultrasound, a more diffusive interaction with the ultrasound pulse is desired. As a result, an 80 μm fiberglass embedded PTFE coating was applied to the instrument in order to improve the appearance.

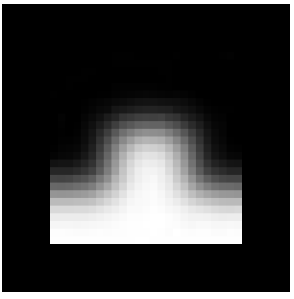


Fig. 5. Passive marker template

Fig. 6. 3D ultrasound image of a surgical instrument in a water tank with white dots indicating tracked passive markers

Finding the markers begins with the image volume already loaded into texture memory from the Radon transform algorithm, then the built-in tri-linear interpolation is used to quickly render a slice through the instrument axis, orientated so that the instrument axis is horizontal. To identify the position of the bumps, a template matching algorithm is used on the ultrasound slice. The algorithm uses the sum of the absolute differences between a candidate region of the slice and a template, shown in Fig. 5.

The positions of the three best matches found in the slice are used to determine the tip position of the instrument and the roll angle (Fig. 6). The tip position is found with a known offset of 3 mm between the two closest markers and the instrument tip. To find the roll angle, the ratio of the distances x_1 and x_2 is used (Fig. 4). Since the third marker is wrapped in a helical pattern around the instrument shaft, the roll angle is a linear function of this ratio.

The configuration of these markers was designed to minimize the computational complexity for real-time implementation. Although there are more compact and higher-resolution marker designs [12], they result in time consuming calculations. In this implementation, the calculations are linear and negligibly add to the computational requirements of the algorithm.

2.3 Experimental Setup

The system uses a Pentium 4 3 GHz personal computer with 1 GB of RAM. The ultrasound machine (SONOS 7500, Philips Medical, Andover, MA) streamed image data over a 1 Gb LAN to a personal computer using TCP/IP. A program written in C++ retrieved the ultrasound volumes and loaded them onto the GPU (7800GT, nVidia Corp, Santa Clara, CA) through a PCI-express bus.

To prove the effectiveness of the complete instrument tracking algorithm, we performed two validation experiments. The first was a water tank study and measured the accuracy of the full six degree of freedom instrument tracking. Six positions on a acetyl block were marked in a 1 cm rectangular pattern. The ultrasound probe and 6 marked positions were held statically, while the

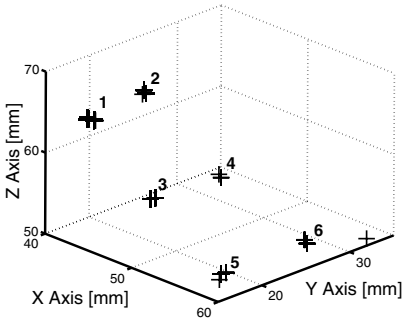


Fig. 7. Plot of the tip location for an instrument in 1 cm spaced positions in a water tank

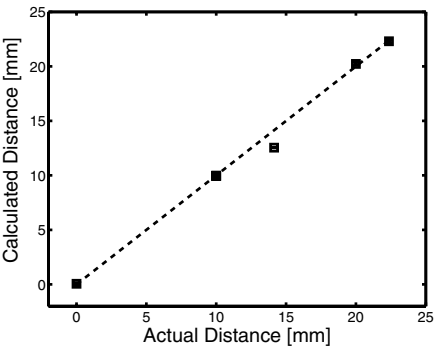


Fig. 8. Distance from position 1 calculated by the tracking algorithm

instrument was moved from position to position. At each of the 6 positions the instrument was imaged for 5 s. The tip position calculated by the tracking algorithm was logged to a data file for analysis.

Second, *in vivo* validation was performed by tracking an instrument within a beating porcine heart. Electromagnetic tracking (miniBIRD 800, Ascension Technology, Burlington, VT) was used to simultaneously track the instrument tip position. The instruments were imaged inside a porcine heart during an open chest beating heart procedure. The instruments were inserted through ports in the right atrial wall and secured by purse-string sutures. The ultrasound probe was positioned epicardially on the right atrium to give a view of the right and left atrium. The surgeon was instructed to move the instrument toward the atrial septum. During this movement, the instrument tip position calculated by the algorithm and the tip position measured by the electromagnetic trackers was recorded to a data file.

3 Results

The algorithm was shown to be effective at real-time tracking surgical instruments in both the water tank and *in vivo* experiments. In the tank experiment, the tracking algorithm accurately measured the relative distances traveled by the instrument tip as it moved to 6 positions 1 cm apart on an acetyl block. Fig. 7 shows the tip positions as reported by the algorithm. The estimated distance of points 2-5 from point 1 is shown in Fig. 8; these distances correspond to the actual separation of the six points marked on the block. The algorithm found distances to points 2 and 3 with a mean error of 0.0 mm and with standard deviation of 0.1 mm. The distance to point 4 showed the largest mean error of 1.6 mm$\pm$0.01 mm and points 5 and 6 had mean errors of 0.2$\pm$0.1 mm and 0.1$\pm$0.1 mm, respectively.

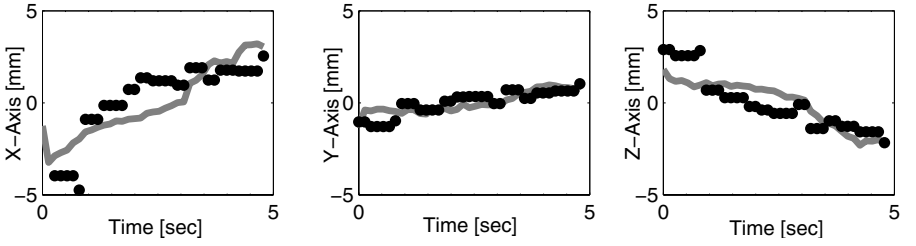


Fig. 9. *In vivo* x, y, and z position of the instrument tip as reported by an electromagnetic tracker (solid line) and the tracking algorithm (dots)

Fig. 9 shows the results of the *in vivo* trials. The figure plots the x, y, and z trajectories of the instrument tip versus time. Both the instrument tip position reported by the instrument tracking algorithm and the electromagnetic trackers is shown. The tracking method correctly tracked the instrument tip as the surgeon moved it for 5 s. The mean difference between the tip position reported by the electromagnetic tracker and tracking algorithm was 1.4 mm, 0.4 mm, and 0.9 mm in x, y, and z directions.

In our experimental setup, the instrument tracking technique required 1.7 s to initially detect the instrument in the entire ultrasound volume. For subsequent tracking, the algorithm required 32 ms per volume. This speed is well within the 38 ms required for the algorithm to keep pace with the 26 volumes per second generated by the ultrasound machine.

4 Discussion

These results show that real-time tracking of surgical instruments is possible with 3DUS. The method successfully integrates Radon transform-based instrument axis tracking and passive marker-based tip and roll tracking to determine the full 6 degrees of freedom of the instrument. In addition, by harnessing the computational power of the GPU and using a fast search schema, the tracking system kept pace with the 25-28 volumes per second produced by the 3DUS machine. To the authors knowledge, this is the first demonstration of real-time tracking of surgical instruments *in vivo* with 3DUS.

The tank experiments highlighted the accuracy of the technique, within 0.2 mm for most cases. However, as seen in one of the measurements (Fig. 8), the marker detection algorithm incorrectly identifies the position of the passive markers. As a result, the tip position is incorrectly calculated along the shaft axis. Further research is underway to improve upon the marker template matching procedure, and to introduce predictive filters to improve accuracy.

In vivo trials provided verification of the effectiveness of the algorithm when instruments are surrounded by blood and highly inhomogeneous and rapidly moving tissue within the beating heart. The electromagnetic tracking used for verification is by no means a "gold standard", as it is highly susceptible to

electromagnetic interference that is found in a modern operating room. It did provide a corroborative data set that showed the tracking technique presented here is in fact tracking the surgical instrument. A more rigorous study into how accurately the method performs *in vivo* is left for future work.

By developing a real-time instrument tracking technique, it is now possible to introduce a range of improvements to aid in our target procedures, intracardiac surgery. Real-time tracking can now be used for instrument overlays and navigational aids to help the surgeon deal with the distorted appearance of tissue and instruments in 3DUS images. In addition, tracking of instruments opens a wide range of possibilities for implementing robot control under 3DUS guidance.

Acknowledgments

This work is supported by the National Institutes of Health under grant NIH R01 HL073647-01.

References

1. Cannon, J.W., et al.: Real time 3-dimensional ultrasound for guiding surgical tasks. Computer Assisted Surgery **8** (2003) 82–90
2. Suematsu, Y., et al.: Three-dimensional echocardiography-guided beating-heart surgery without cardiopulmonary bypass: a feasibility study. J Thorac Cardiovasc Surg. **128** (2004) 579–587
3. Suematsu, Y., et al.: Three-dimensional echo-guided beating heart surgery without cardiopulmonary bypass: atrial septal defect closure in a swine model. J Thorac Cardiovasc Surg. **130** (2005) 1348–57
4. Murkin, J.M., et al.: Beating heart surgery: Why expect less central nervous system morbidity? Annals of Thoracic Surgery **68** (1999) 1498–1501
5. Zeitlhofer, J., et al.: Central nervous system function after cardiopulmonary bypass. European Heart Journal **14** (1993) 885–890
6. Bellinger, D., et al.: Developmental and neurological status of children at 4 years of age after heart surgery with hypothermic circulatory arrest or low-flow cardiopulmonary bypass. Circulation **100** (1999) 526–532
7. Vitrani, M., Morel, G., Ortmaier, T.: Automatic guidance of a surgical instrument with ultrasound based visual servoing. Proc. IEEE ICRA (2005) 510–515
8. Ortmaier, T., et al.: Robust real-time instrument tracking in ultrasound images for visual servoing. Proc. IEEE ICRA (2005) 2167–72
9. Draper, K., et al.: An algorithm for automatic needle localization in ultrasound-guided breast biopsies. Medical Physics **27** (2000) 1971–9
10. Stoll, J., Novotny, P., Dupont, P., Howe, R.: Real-time 3d ultrasound-based servoing of a surgical instrument. Proc. IEEE ICRA (2006)
11. Novotny, P., Zickler, T., Howe, R.: Radon transform based instrument shaft detection in three-dimensional ultrasound. submitted (biorobotics.harvard.edu/pubs/NovotnyTMI.pdf) (2006)
12. Stoll, J., Dupont, P.: Passive markers for ultrasound tracking of surgical instruments. MICCAI (2005)

Shape-Driven 3D Segmentation Using Spherical Wavelets

Delphine Nain[1], Steven Haker[2], Aaron Bobick[1], and Allen Tannenbaum[3]

[1] College of Computing, Georgia Institute of Technology, Atlanta, USA
[2] Department of Radiology, Brigham and Women's Hospital, Boston, USA
[3] Electrical Engineering, Georgia Institute of Technology, Atlanta, USA

Abstract. This paper presents a novel active surface segmentation algorithm using a multiscale shape representation and prior. We define a parametric model of a surface using spherical wavelet functions and learn a prior probability distribution over the wavelet coefficients to model shape variations at different scales and spatial locations in a training set. Based on this representation, we derive a parametric active surface evolution using the multiscale prior coefficients as parameters for our optimization procedure to naturally include the prior in the segmentation framework. Additionally, the optimization method can be applied in a coarse-to-fine manner. We apply our algorithm to the segmentation of brain caudate nucleus, of interest in the study of schizophrenia. Our validation shows our algorithm is computationally efficient and outperforms the Active Shape Model algorithm by capturing finer shape details.

1 Introduction

The characterization of local variations in a shape population is an important problem in medical imaging since a disease could affect only a portion of an organs surface. For example, the surfaces of certain structures in the brain, such as the caudate nucleus, contain sharp features. During the segmentation process of such structures, this relevant high frequency information needs to be preserved since it is relevant to further shape analysis [1].

Object segmentation with deformable models and statistical shape modelling are often combined to obtain a robust and accurate segmentation [2, 3, 4, 5]. Active shape models (ASMs) [3] are a standard technique for statistical segmentation tasks based on a prior learned over a point distribution models (PDM): landmarks on the shape are used as parameters and a joint prior probability distribution is learned using principal component analysis (PCA) over the landmarks. However, ASMs are often limited by the training set size and the inability of relatively few eigenvectors to capture the full biological variations in the data [6].

To address this, a decomposable shape representation seems natural, where shape descriptors are separated into groups that describe independent global and/or local biological variations in the data, and a prior induced over each group explicitly encodes these variations. Wavelet basis functions are useful for such a representation since they range from functions with global support to functions localized both in frequency and space, so that their coefficients can

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 66–74, 2006.

be used both as global and local shape descriptors, unlike spherical harmonics or PCAs over landmarks which are global shape descriptors. The authors in [6] used wavelet functions for curve representation and learned a prior over groups of coefficients that are in proximity both in scale and space used for segmentation of 2D medical imagery, with convincing results. In [7], authors presented a multiscale representation of 3D surfaces using conformal mappings and spherical wavelets, and a novel multiscale shape prior shown to encode more descriptive and localized shape variations than the PDM shape prior for a given training set size. In this work, we present a segmentation framework using this 3D wavelet representation and multiscale prior. To the best of our knowledge, this is the first application of spherical wavelets for medical image segmentation.

In Section 2, we give an overview of the shape representation and shape prior using spherical wavelets [7]. Both will be used in the segmentation framework described in Section 3. In Section 4, we present our results on caudate datasets, and in Section 5 we summarize our results and outline further research.

2 Shape Representation and Prior

2.1 Spherical Wavelets

In this work, we use biorthogonal spherical wavelets functions described in [8]. Spherical wavelets are scalar functions defined on surfaces which are topologically equivalent to the unit sphere and equipped with a multiresolution mesh, created by recursively subdividing an initial mesh so that each triangle is split into 4 "child" triangles at each new subdivision (resolution) level (see Figure 1). At each resolution level, scaling and wavelet functions are constructed, with decreasing support as the resolution increases (see Figure 1). For a mesh with N vertices, a total of N functions constitute an L^2 basis, which means that any finite energy function defined on the mesh can be expressed in the basis. In matrix form, the set of basis functions can be stacked as columns of a matrix Φ of size $N \times N$ where each column is a basis function evaluated at each of the N vertices. Since the spherical wavelet functions are biorthogonal, $\Phi^T \Phi \neq Id$ (the identity matrix), so the inverse basis Φ^{-1} is used for perfect reconstruction, since $\Phi^{-1}\Phi = Id$.

Any finite energy scalar function evaluated at N vertices, denoted by the vector F of size $N \times 1$, can be transformed into a vector of basis coefficients Γ_F of size $N \times 1$ using the **Forward Wavelet Transform**: $\Gamma_F = \Phi^{-1} F$, and recovered using the **Inverse Wavelet Transform**: $F = \Phi \Gamma_F$.

Next, we describe how to represent shapes using spherical wavelets.

2.2 Data Description

To illustrate our work, we use a dataset of 29 left caudate nucleus[1]. The MRI scans were hand-segmented by an expert neuroanatomist to provide ground truth segmentations. Each manual segmentation defined a 3D surface extracted by standard isosurface algorithm. We used 24 training shapes and 5 test shapes.

[1] The details are: 1.5 Tesla GE Echospeed system, coronal SPGR images, 124 slices of 1.5 mm thickness, voxel dimensions $0.9375 \times 0.9375 \times 1.5mm$.

2.3 Shape Representation

We first re-triangulate and register surfaces in the dataset so that they each have the required multiresolution mesh to conduct the wavelet analysis and the same mesh node on all shapes correspond to the same anatomical location. To achieve this, we find a one-to-one mapping from each surface of the population to the sphere, constrained by the requirements that 6 consistently chosen landmark points[2] on the shapes get mapped to the same point on the sphere [10]. Then, by interpolating the resulting spherical maps at the vertices of a multiresolution triangulation of the sphere (shown in Figure 1(b)), we can re-triangulate the original caudate surfaces in a consistent manner, providing a point-by-point registration of all surfaces in the dataset and the required mesh for spherical wavelet analysis. After registration, all shapes have N vertices and the i^{th} shape is represented by the vector Σ_i of size $3N$ (all x, then y then z coordinates). All shapes are aligned with Procrustes [11] to find the mean shape $\overline{\Sigma}$.

We encode the deviation from the mean for the i^{th} shape with the signal $v_i = \Sigma_i - \overline{\Sigma}$. We then transform v_i into a vector of spherical wavelet basis coefficients Γ_{v_i} of size $3N$ with the forward spherical wavelet transform:

$$\Gamma_{v_i} = \underbrace{\begin{bmatrix} \Phi^{-1} & 0 & 0 \\ 0 & \Phi^{-1} & 0 \\ 0 & 0 & \Phi^{-1} \end{bmatrix}}_{\Pi^{-1}} v_i, \tag{1}$$

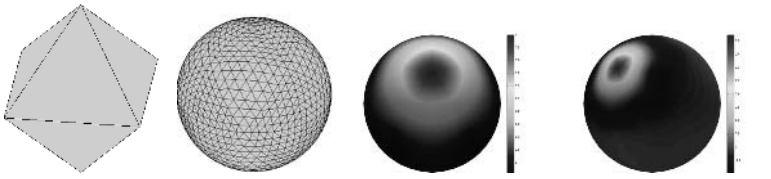
Therefore a shape is transformed into wavelet coefficients by taking the forward wavelet transform of the x, y and z deviation from the mean signal.

2.4 Multiscale Shape Prior

To build a prior that captures both global and local variations in the data, we first reduce the dimensionality of the Γ_{v_i} coefficients and keep only the coefficients that encode relevant variations in the training data. We use a technique that takes into account biorthogonality and estimates which coefficients can be truncated (set to 0) without significantly affecting the function approximation [9]. In the caudate dataset, 74% of the coefficients were removed resulting in a reconstruction error smaller than 0.1% of the total shape size. This leads to a nice compression property since the transform can match variations in the caudate shape population using a small number of basis functions.

After truncation, we wish to decompose the set of non-truncated coefficients (vector of size $3M$) such that *highly correlated* coefficients are grouped together in a band, with the constraint that coefficients across bands have minimum cross-correlation as described in [7]. This models the joint probability distribution of the coefficients by a product of smaller probability distributions over each band, assumed to model independent shape variations at a particular scale. We recursively cluster coefficients using spectral graph partitioning [12]. The

[2] The landmarks are chosen automatically as described in [9].



(a) level 0 (b) level 4 (c) wavelet, level 1 (d) wavelet, level 2

Fig. 1. (a-b) Recursive Partitioning of an octahedron (c-d) Visualization of wavelet basis functions at various levels. The color corresponds to the value of the functions.

(a) Band 1 at scale 1, Lateral View 1 (b) Band 1 at scale 1, Lateral View 2 (c) Band 2 at scale 1, Lateral View 1 (d) Band 2 at scale 1, Lateral View 2

(e) Band 3 at scale 2, Lateral View 1 (f) Band 3 at scale 2, Lateral View 2 (g) Color Scale

Fig. 2. 3 examples bands discovered by the prior color-coded on the mean shape. The color shows the cumulative value of the wavelet basis functions that belong to that band. Whiter areas represent surface locations with correlated variations across shapes.

visualization of resulting bands on the mean shape can in itself be interesting for shape analysis (see Figure 2) by indicating which surface patches co-vary across the training set. For example at scale 1, bands 1 and 2 indicate two uncorrelated shape processes in the caudate data that make sense anatomically: the variation of the head and of the body. It is interesting that bands have compact spatial support, though this is not a constraint of our technique.

Once we discovered the bands, the final step is to estimate the probability distribution of each band of coefficients via PCA. The eigenvectors and eigenvalues of lower scale bands represent relatively global aspects of shape variability, whereas bands at higher scales represent higher frequency and more localized aspects of shape variability. Additionally, our prior accurately encodes finer details even with small training sets, since if there are a total of B bands, there exists on the order of $L \approx B(K-1)$ eigenvectors, as opposed to just $K-1$ eigenvectors when performing PCA on PDMs.

The full prior contains all the eigenvectors for all bands and all resolutions in a matrix U of size $3M \times L$ if there are L eigenvectors in total[3]. The vector of basis coefficients Γ_{v_i} is then :

$$\Gamma_{v_i} = \overline{\Gamma_{v_i}} + U\alpha_{v_i} \tag{2}$$

[3] Each column of U is an eigenvector of a particular band with non-zero entries only for coefficients that were assigned to that band, see [7, 9] for more details.

where α_{v_i} (size $L \times 1$) represents the coordinates of the wavelet coefficients of that shape in the eigenvector space.

3 Segmentation with the Spherical Wavelet Prior

In order to exploit the multiscale prior, we derive a parametric surface evolution equation by *evolving the weights* α directly. As the surface evolves to fit the image data, we constrain the weights α to remain within ± 3 standard deviation of their values observed in the training set. The parameters of our model are the shape parameters α, as well as pose parameters that accommodate for shape variability due to a similarity transformation (rotation, scale, translation) which is not explicitly modelled with the shape parameters.

3.1 Pose Parameters

Given a surface mesh with N vertices $\Sigma : [1, ..., N] \rightarrow \mathbb{R}^4$, expressed in homogeneous coordinates so that a mesh point is denoted by $\Sigma(u) = \mathbf{x}_u = [x_u, y_u, z_u, 1]^T$, a transformed surface $\widetilde{\Sigma}$ is defined by:

$$\widetilde{\Sigma}(u) = T[\mathbf{p}]\Sigma(u). \tag{3}$$

The transformation matrix $T[\mathbf{p}]$ is the product of a translation matrix with 3 parameters t_x, t_y, t_z, a scaling matrix with 1 parameter s and a rotation matrix with 3 parameters w_x, w_y, w_z, using the exponential map formulation [13].

3.2 Shape Parameters

A surface point $\Sigma(u)$ can be represented in the wavelet basis using (1) and (2):

$$\Sigma(u) = \overline{\Sigma}(u) + \mathcal{H}\left(\Pi_u \Gamma\right) = \overline{\Sigma}(u) + \mathcal{H}\left(\Pi_u(\overline{\Gamma} + U\alpha)\right) \tag{4}$$

where the function $\mathcal{H} : [3N \times 1] \rightarrow [4 \times N]$ rearranges a matrix to have correct homogeneous coordinates and Π_u are all the basis functions in Π evaluated at point $\mathbf{x}_u$. The parameters α are the shape parameters of our model.

3.3 Segmentation Energy

We use a region-based energy to drive the evolution of the parametric deformable surface for segmentation. With region-based energies, the force that influences the evolution of a contour depends on more global statistical information [4, 5]. We employ the discrete version of a segmentation energy presented in [5]:

$$E(\alpha, \mathbf{p}) := \sum_{\widetilde{\mathbf{x}} \in \widetilde{R}} L(\widetilde{\mathbf{x}}) \Delta \widetilde{\mathbf{x}}, \tag{5}$$

where $\widetilde{R}$ is the region inside the evolving surface $\widetilde{\Sigma}$ and the force is $L(\widetilde{\mathbf{x}}) = -\log(\frac{P_I(I(\widetilde{\mathbf{x}}))}{P_O(I(\widetilde{\mathbf{x}}))})$ where $I(\widetilde{\mathbf{x}})$ is the image intensity at a point $\widetilde{\mathbf{x}}$ located inside the region $\widetilde{R}$ of the evolving surface, $P_I(I(\widetilde{\mathbf{x}}))$ is the probability that a point $\widetilde{\mathbf{x}}$ with intensity $I(\widetilde{\mathbf{x}})$ belongs to the interior of an object to be segmented in the image,

and P_O is the probability that the point belongs to the exterior of the object. The segmentation energy is minimized when the surface evolves to include points that have maximum L (points that have a higher P_I than P_O). To estimate the probability density functions P_I and P_O from a training set, we collect sample voxel intensity values inside and outside the segmented shapes in a neighborhood of width 10 pixels around the boundary and use Parzen windows [11].

The surface evolution is defined by a gradient of Σ that minimizes the energy *in terms of the pose* $\mathbf{p}$ *and shape parameters* α. We use the area formula, and discrete divergence theorem to express the region sum in (5) as a surface sum [9]. Using the notation of (4), the gradient with respect to each pose parameter $p_k \in \mathbf{p}$ is given by:

$$\frac{dE}{dp_k} = \sum_{\widetilde{\mathbf{x}_u} \in \widetilde{\Sigma}} < L(\widetilde{\mathbf{x}_u}) \frac{dT[\mathbf{p}]}{dp_k} \Sigma(u), \widetilde{\mathcal{N}} > \Delta\widetilde{\mathbf{x}_u}, \tag{6}$$

$\widetilde{\mathcal{N}}$ is the inward normal of surface point $\widetilde{\mathbf{x}_u}$ expressed in homogeneous coordinates and the image force L is evaluated at points on the surface boundary of $\widetilde{\Sigma}$.

The gradient flow with respect to each shape parameters $\alpha_k \in \alpha$ is given by:

$$\frac{dE}{d\alpha_k} = \sum_{\widetilde{\mathbf{x}_u} \in \widetilde{\Sigma}} < L(\widetilde{\mathbf{x}_u}) T[\mathbf{p}] \mathcal{H}(\Pi_u U(:,k)), \widetilde{\mathcal{N}} > \Delta\widetilde{\mathbf{x}_u} \tag{7}$$

where $U(:,k)$ selects the k^{th} eigenvector corresponding to α_k.

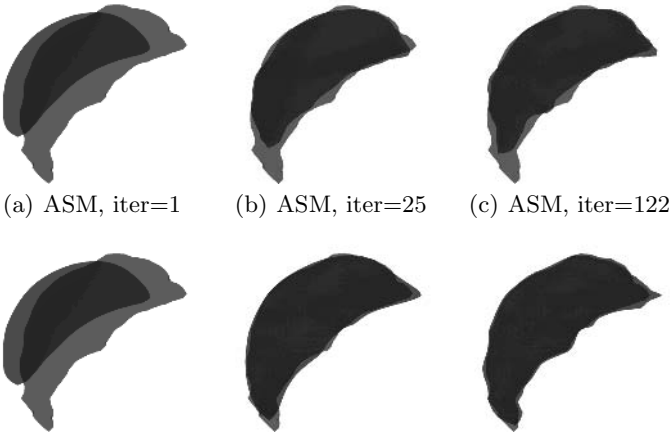
3.4 Parameter Optimization Via Multiresolution Gradient Descent

We can now use the gradient equations (6) and (7) to conduct a parameter optimization via gradient descent. Explicitly, the update equations are:

$$\mathbf{p}(t+1) = \mathbf{p}(t) + \delta_t^{\mathbf{p}} \frac{dE}{d\mathbf{p}} \quad (8) \qquad \alpha(t+1) = \alpha(t) + \delta_t^{\alpha} \frac{dE}{d\alpha} \quad (9)$$

where δ_t^{α} and $\delta_t^{\mathbf{P}}$ are positive step size parameters and $\alpha(t+1)$, $\mathbf{p}(t+1)$ denote the values of the parameters α and $\mathbf{p}$ at the $(t+1)^{th}$ iteration. We start with an initial shape and iterate between (8) and (9). We update the α parameters in a *multiresolution* fashion. Since each shape parameter α_i corresponds to a band at a wavelet resolution j, we first only update α coefficients corresponding to the coarsest level bands ($j = 1$). Once α changes less than a threshold value v_α, we add the α parameters of the next resolution level to the gradient and update (9). This results in a more stable segmentation since few global parameters are first updated when the shape is far from the solution, and more localized parameters are added as the shape converges to the solution.

We start with (8) until $(\mathbf{p}^{t+1} - \mathbf{p}^t) < v_{\mathbf{p}}$ where $v_{\mathbf{p}}$ is a threshold value. We then run (9) for 1 iteration, and iterate the process. At each α iteration, we ensure that the value of the α parameters stays within ± 3 standard deviation of the observed values in the training set. After each iteration, the updated shape and pose parameters are used to determine the updated surface.



(a) ASM, iter=1 (b) ASM, iter=25 (c) ASM, iter=122

(d) Mscale, iter=1 (e) Mscale, iter=96, levels 1-2 active (f) Mscale, iter=122, levels 1-5 active

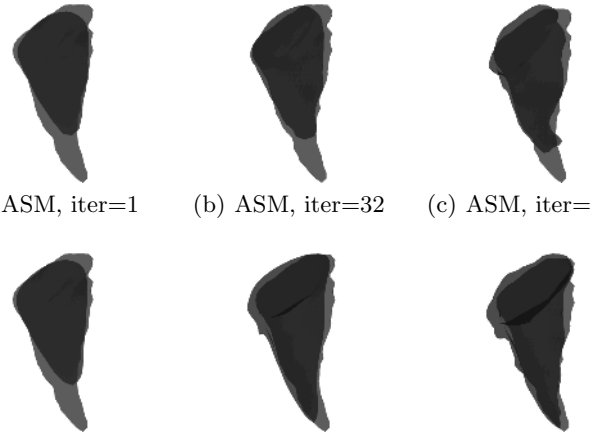
Fig. 3. Surface Evolution using the **Ground Truth** labelmap as the image force for ASM (top rown) and Mscale (bottom row) algorithms. The ground truth is shown in light gray, the evolving surface in dark grey.

4 Results

We have applied our algorithm to the segmentation of caudate nucleus shapes from MRI scans as previously described. We learned a shape prior from a training set of 24 shapes. We use spherical wavelet basis functions of resolution up to $j = 5$. In total, we obtained 16 bands in the shape prior. We learned the mean position p_m of the caudate shapes in the MRI scans (in patient coordinate space). To initialize the segmentation, we use the mean caudate shape learned during the training phase and positioned it at position p_m in the scan to be segmented. We then evolved the surface according to the process described in Section 3. The step size parameters were $\delta_t^\alpha = 0.5, \delta_t^{\mathbf{P}} = 0.001$ for translation and $\delta_t^{\mathbf{P}} = 0.0001$ for scale and rotation and $v_{\mathbf{p}} = v_\alpha = 0.02$.

To measure the discrepancy between the segmented shape (S) and the ground truth (G) (obtained from the hand-segmented labelmaps), we use the Hausdorff distance $H(G, S)$ that measures the maximum error between the boundary of two shapes G and S, as well as the partial Hausdorff distance $H_f(G, S)$ that measures the $f\%$ percentile of the Hausdorf distance. We compare our algorithm (called Mscale) to the standard ASM algorithm that uses the PDM shape prior [3], using the same training, testing shapes and keep 100% of the eigenvectors.

To validate our algorithm, we first use the Ground Truth labelmap as the image force in Equations (6, 7) by replacing the log expression with a value of 1 inside the (known) object and -1 outside. The end goal is to validate whether the surface evolution converges to the right solution, given perfect image information. Since we are evolving in the space of the shape prior, the discrepancy between the ASM and Mscale algorithm is due to the expressiveness of the shape prior. Figure 3 shows the result for test shape 5. The final segmentation with the



(a) ASM, iter=1 (b) ASM, iter=32 (c) ASM, iter=202

(d) Mscale, iter=1 (e) Mscale, iter=150, levels 1-2 active (f) Mscale, iter=202, levels 1-5 active

Fig. 4. Surface Evolution using the **density estimation** as the image force for ASM (top rown) and Mscale (bottom row) algorithms. The ground truth is shown in light gray, the evolving surface in dark grey.

Table 1. Hausdorff (H) and partial Hausdorff (H_f) distance for the five test shapes, for the Multiscale (Mscale) and Active Shape Model segmentation algorithm.

Error Measure	Segmentation Alg.	N=1	N=2	N=3	N=4	N=5	Mean
H_{95} (mm)	Mscale	**4.82**	**2.22**	**3.03**	**3.04**	**3.95**	**3.16**
	ASM	5.51	3.24	3.98	3.18	4.26	3.83
H (mm)	Mscale	**5.89**	**4.06**	**3.75**	**5.23**	**5.57**	**4.85**
	ASM	9.79	5.68	6.33	7.22	6.06	7.07

multiscale prior captures more of the shape and finer details than the ASM segmentation. Furthermore, we see that as the resolution level is increased for the α parameters, the Mscale segmentation is able to capture finer details.

We then validated the full segmentation algorithm, using the proposed image force in Equation 5. The results of the validation for both algorithms are shown in Table 1. For each test shape, the lowest error among the two algorithms is in boldface. The Mscale algorithm consistently outperforms the ASM algorithm. Figure 4 qualitatively compares the segmentation of Test shape 3 for both algorithms. The Mscale algorithm is more accurate and captures finer details, especially at the tail of the shape. We note that the segmentation is not fully accurate due to non-perfect image statistics. Our algorithm runs under 5 minutes on a Pentium IV 2GHz using non-optimized MATLAB code.

5 Conclusions and Future Work

We presented a computationally efficient segmentation algorithm based on a spherical wavelet shape representation and multiscale shape prior. Our results

show that the proposed segmentation algorithm outperforms standard ASM by capturing finer details during the evolution, due to the expressiveness of the multiscale prior that captures a wider range of global and local variations in the training set than the ASMs prior. We plan to validate the algorithm on other medical structures, in particular other brain structures involved in diagnosis of schizophrenia. Additionally, we plan to investigate the usefulness of the spherical wavelet shape representation and multiscale prior for shape classification.

Acknowledgements. This work is part of the National Alliance for Medical Image Computing (NAMIC), funded by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant U54 EB005149. The work of A. Tannenbaum and D. Nain is also supported by NIH grant NAC P41 RR-13218 as well as a Marie Curie Grant from the EU through the Technion. S. Haker's work is supported by NIH grants R01CA109246, U41RR019703, P41RR013218, R01CA104976, R01LM007861.

References

1. S.Vetsa, Styner, M., Pizer, S., Lieberman, J., Gerig, G.: Caudate shape discrimination in schizophrenia using template-free non-parametric tests. In: Medical Image Computing and Computer-Assisted Intervention MICCAI. (2003) 661–669
2. Xu, C., Pham, D.L., Prince, J.L.: Medical Image Segmentation Using Deformable Models. In: Handbook of Medical Imaging, Volume 2. SPIE Press (2000) 129–174
3. Cootes, T.F., Taylor, C.J.: Combining point distribution models with shape models based on finite element analysis. Image Vis. Computing **13** (1995) 4039
4. Tsai, A., Yezzi, A.J., III, W.M.W., Tempany, C., Tucker, D., Fan, A., Grimson, W.E.L., Willsky, A.S.: A shape-based approach to the segmentation of medical imagery using level sets. IEEE Trans. Med. Imaging **22** (2003) 137–154
5. Rousson, M., Cremers, D.: Efficient kernel density estimation of shape and intensity priors for level set segmentation. In: MICCAI. (2005)
6. Davatzikos, C., Tao, X., Shen, D.: Hierarchical active shape models, using the wavelet transform. IEEE Trans. Medical Imaging **22** (2003) 414–423
7. Nain, D., Haker, S., Bobick, A., Tannenbaum, A.: Multiscale 3d shape analysis using spherical wavelets. MICCAI (2005)
8. Schröder, P., Sweldens, W.: Spherical wavelets: Texture processing. In Hanrahan, P., Purgathofer, W., eds.: Rendering Techniques '95. Springer Verlag (1995)
9. Nain, D., Haker, S., Bobick, A., Tannenbaum, A.: Multiscale 3d shape representation and segmentation using spherical wavelets. Submitted journal article (2006)
10. Haker, S., Warfield, S., Tempany, C.: Landmark-guided surface matching and volumetric warping for improved prostate biopsy targeting. MICCAI (2004)
11. Duda, R., Hart, P., Stork, D.: Pattern Classification. Wiley-Interscience (2001)
12. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (2000) 888–905
13. Lepetit, V., Fua, P.: Monocular Model-Based 3D Tracking of Rigid Objects: A Survey. In: Foundations and Trends in Computer Graphics and Vision. Volume 1. Now Publishing (2005) 1–89

Artificially Enlarged Training Set in Image Segmentation

Tuomas Tölli[1], Juha Koikkalainen[2], Kirsi Lauerma[3], and Jyrki Lötjönen[2]

[1] Laboratory of Biomedical Engineering, Helsinki University of Technology, P.O.B. 2200, FIN-02015 HUT, Finland
tuomas.tolli@tkk.fi
[2] VTT Information Technology, P.O.B. 1206, FIN-33101 Tampere, Finland
{juha.koikkalainen, jyrki.lotjonen}@vtt.fi
[3] Helsinki Medical Imaging Center, University of Helsinki, P.O.B. 281, FIN-00029 HUS, Finland

Abstract. Due to small training sets, statistical shape models constrain often too much the deformation in medical image segmentation. Hence, an artificial enlargement of the training set has been proposed as a solution for the problem. In this paper, the error sources in the statistical shape model based segmentation were analyzed and the optimization processes were improved. The method was evaluated with 3D cardiac MR volume data. The enlargement method based on non-rigid movement produced good results – with 250 artificial modes, the average error for four-chamber model was 2.11 mm when evaluated using 25 subjects.

1 Introduction

Quantitative information, such as various volumetric measures, are needed to make more objective diagnosis. However, manual extraction of such information from images is both highly time-consuming and prone to segmentation errors due to subjective image interpretation. Therefore, accurate and reliable methods to automatically perform the segmentations are highly needed. Development of automatic segmentation methods has appeared to be challenging due to several reasons, such as high variability in the shape and appearance of various anatomical regions, low contrast between objects and noise in images.

Numerous model-based segmentation methods have been studied to solve the aforementioned difficulties. Especially active shape models (ASM) [1] has turned out to give promising results. In ASM, the mean model and its typical deformation modes, based on a training set, are defined using Principal Component Analysis (PCA). Several ASM-based approaches have been proposed for the 3D modeling and segmentation of medical images [2, 3, 4, 5]. Furthermore, the methods have lately been used to model the function of the heart [6].

Because the model construction process is time consuming and not enough data are always available, the number of examples in the training set remains often relatively low. This means that the degrees of freedom in the model are limited and the model can not represent accurately all cases in a population from

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 75–82, 2006.

which the training set has been sampled. At least two techniques have been proposed to the problem: for example in [7, 8], the approach was to decrease the dimension of the model, whereas in [9, 10], the size of the training set was artificially increased.

In addition to testing different methods to enlarge the dataset artificially in [10], we studied the relation between the size of the training set and the ability of the model to represent an unseen object. We found that a heart model with 100 modes has enough degrees of freedom to reach about 1 mm segmentation accuracy for ventricles and atria. However, preliminary tests with real data showed that the segmentation accuracy did not improve as much as the modeling accuracy.

In this work, we identified three sources for the segmentation error: 1) the properties of the model, i.e., degrees of freedom in the model, 2) the properties of the deformation algorithm, such as goodness of criterion to be optimized and an optimization algorithm used, and 3) the errors in the golden standard to which a segmentation result is compared. Our objective is to study the contribution of these error sources to the final segmentation result. This was obtained 1) by proposing a method to eliminate the effects of the manual segmentation errors on the segmentation accuracy, 2) by improving and testing different optimization methods to deform the shape model and 3) by further analyzing and verifying the effects of the artificial enlargement of the training set. Image data used consisted of cardiac MR volumes.

Another contribution made in this work was the development of a novel segmentation algorithm for cardiac MR images. The method was based on artificially enlarged ASM. The weights of the deformation modes were optimized by fitting line profiles, defined for model points, to image data by simulated annealing.

2 Materials and Methods

Materials. The database consisted of both long- and short-axis MR volumes obtained from 25 healthy subjects. Atria, ventricles and epicardium were manually segmented from the volumes by an expert. The segmentations were made by fitting a triangulated surface model simultaneously into the short- and long-axis volumes. All manual segmentations were made using a specially designed software package [3].

Model Construction. Both the mean shape $\bar{\mathbf{s}}$ and the mean gray-scale short- and long-axis volumes were constructed according to the procedure proposed in [3]. Also, the point correspondence between the training set shapes $\mathbf{s}_i$ were obtained, and used to model the shape variability in the training set using point distribution model (PDM) [1]. A covariance matrix was computed from the deviations of the training set shapes from the mean shape. The deformation modes, i.e. modes of shape variation, were defined by computing eigenvectors and -values of the covariance matrix.

In statistical shape models, new examples of the shape, $\mathbf{s} = [s_1, \ldots, s_n]^T$, that are specific to the studied object, are generated using a linear combination

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{\Phi b} \ , \tag{1}$$

where $\bar{\mathbf{s}} = [\bar{s}_1, \ldots, \bar{s}_n]^T$ is a reference shape, typically a mean shape constructed from a training set, $\mathbf{\Phi} = [\phi_1, \ldots, \phi_m]$ is a matrix consisting of the deformation modes ϕ_i, and $\mathbf{b} = [b_1, \ldots, b_m]^T$ is a weight vector. The number of the non-zero eigenvectors is $min(N-1, n)$, where N is the number of the training set shapes and n the number of shape parameters, i.e. point coordinates in PDM.

To avoid the training set constricting too much the model deformation during the segmentation process, the size of the training set, and hence, the number of the deformation modes was artificially enlarged. In [10], we evaluated several methods and the best one called a non-rigid movement strategy was adopted here. The method moves non-rigidly surface points inside a deformation sphere. The sphere is randomly positioned to 50 locations on the object surfaces, and the original surface points are locally deformed at each position. The displacement vector for the movement is calculated from

$$\mathbf{v}(x, y, z) = \frac{e^{-2\frac{(x-c_x)^2+(y-c_y)^2+(z-c_z)^2}{r^2}} - e^{-2}}{1.0 - e^{-2}}\mathbf{V} \ , \tag{2}$$

where (c_x, c_y, c_z) and r are the location and the radius of the sphere ($r = 50$ mm), and $\mathbf{V}$ a random vector. The length of the vector $\mathbf{V}$ is chosen from a uniform distribution ([0 25] mm). Outside the sphere, the displacement vector is zero.

In ASM-based segmentation, the mean model is deformed according to Eq. 1 in such a way that the deformed model fits optimally to the target data. In other words, the objective is to find a combination of weights b_i optimizing a selected similarity function. In this work, the weights were constricted to satisfy $|b_i| \leq 5\sqrt{\lambda_i}$, in which λ_i is the eigenvalue corresponding to the ith eigenvector.

Optimization Criteria. Two optimization criteria were tested for the model deformation. Both criteria were based on the use of line profiles normal to the subject surface [11]. The first method utilized Normalized Mutual Information (NMI) [12]. For each surface point j, the grayscale information along the profile was sampled and put into the vector $\mathbf{g}_j$. The varying of the weight vector $\mathbf{b}$ was made in such a way that the NMI between the profiles from the mean model and the profiles from the same locations in the target data was maximized.

In the second method, the normalized profiles $\mathbf{g}_j$ were built for each surface point of each subject in the training set. Then, their mean $\bar{\mathbf{g}}_j$ and covariance $\mathbf{S}_g$ were calculated. Here, it was assumed that the normalized samples were distributed as a multivariate Gaussian. Thereafter, the quality of fit between the target profiles and the profiles from the mean model was computed by minimizing the sum of the statistical (Mahalanobis) distances, $f(\mathbf{g})$, calculated over all the surface points as

$$f(\mathbf{g}) = \sum_{j=1}^{n} (\mathbf{g}_j - \bar{\mathbf{g}})^T \mathbf{S}_g^{-1} (\mathbf{g}_j - \bar{\mathbf{g}}) \ . \tag{3}$$

Optimization Algorithms. A commonly known problem with deformable models is that the deformation fails easily if the model is not close enough the final result. In other words, the deformation stops to some local minimum/maximum of the optimization function. Therefore, we evaluated in this work both local and global optimization methods. Conjugate gradient and simplex optimization methods were used as local optimization methods. Simulated annealing (SA), a method simulating physical annealing process was used for global optimization. In SA, annealing schedule for the simulated temperature was calculated from

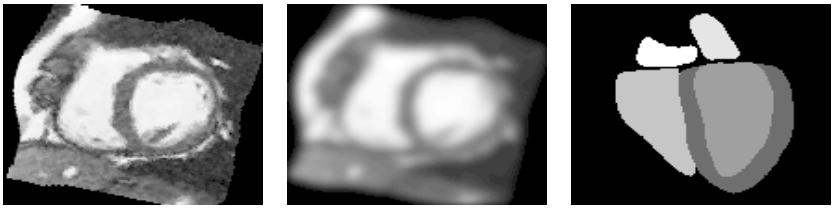
$$T = T_0(1 - \frac{k}{K})^{\alpha} \ , \tag{4}$$

where T_0 is the initial temperature, k the iteration number and K the number of total iterations ($\alpha = 4$). All the implementations pertaining to the optimization methods were based on the algorithms proposed in [13].

Modified Training Set. The segmentations were first done for original MR volumes with both normal PCA modes and artificially enlarged modes, using the mean model as an *a priori* model. However, the errors in the manual segmentations were discovered to have an influence on the process both during the deformation step (wrong shapes in the training set) and during the method validation step (inaccurate golden standard). Because no realistic cardiac MR phantom images were available, the following strategy was used to eliminate the errors in the golden standard.

A modified training set was built by exploiting the transformation fields obtained during the model construction phase. New *a priori* model was obtained by registering the database subject k with the mean model using the volumetric transformation $\mathbf{T}_k : x_k \mapsto \bar{x}$ (Fig. 1a). To better simulate the mean model, the new model was filtered with a Gaussian lowpass filter (Fig. 1b). Moreover, the target volumes were obtained by registering database subject k to all other database subjects l using the transformations $\mathbf{T}_{kl} : x_k \mapsto x_l$. This yielded subjects having the grayscale information from subject k and shapes from subject l. The segmentations were then applied to these volumes using the same statistical shape models than in normal MR volume segmentations. Thus, the manual segmentation errors were propagated both to the mean model and to the database subjects in a similar way, removing the effects on the measured segmentation accuracy.

Evaluation. First, the generality of the statistical shape model based segmentation was verified by applying the method to labeled volumes, where each object was presented by one grayscale value (Fig. 1c). Thereafter, the segmentation was evaluated both with 25 real and modified short- and long- axis MR volumes. In the segmentations, both the NMI- and Mahalanobis-based optimization criteria were used with different optimization algorithms. In addition to standard PCA,



(a) (b) (c)

Fig. 1. An *a priori* model of the modified training set, i.e., one database subject transferred into the mean coordinate system a) before and b) after the Gaussian lowpass filtering process. c) An example of labeled volumes used to verify the generality of the statistical shape model based segmentation.

the segmentations were made using 24, 100 and 250 artificial modes, respectively. During all tests, leave-one-out cross-validation was used to avoid biasing.

Euclidean distances between the results obtained from the automatic and manual segmentation processes were used as error measures. The distances were averaged over all surface points of all subjects. Moreover, comparison between the absolute energy values were also made to better understand the phenomena.

3 Results

The segmentation results are summarized in Table 1. Moreover, the error distributions between different cardiac objects using simulated annealing and Mahalanobis distance based optimization are shown in Table 2. As an example, two randomly selected results are illustrated in Fig. 2, one in short- and one in long-axis view.

4 Discussion

As shown in Table 1, the performance of the method with labeled volumes follows the accuracy estimated in [10], i.e. the modeling accuracy in terms of the number of deformation modes. For example with 100 modes, approximately 1 mm error was produced since the deformation modes did not have more degrees of freedom to represent the full range of shape variations of the target object. Furthermore, the error increased to 1.36 mm when the modified volumes were used. This means that image characteristics, such as image noise and poorly visible interfaces between neighboring objects, increased error approximately 0.3-0.4 mm. This 1.4 mm can be considered as the real segmentation error of the algorithm. Finally, as the errors from the manual segmentations were also present, the error further increased by 0.5-1 mm, being 2.16 mm for real MR volumes using Mahalanobis distance based optimization.

For comparison, we segmented manually the left and right ventricle as well as epicardium from seven cases twice. The difference between these segmentations

Table 1. ***Top:*** Segmentation error and its standard deviation (mm) for simulated annealing. The standard PCA as well as the PCA with 24, 100 and 250 artificial modes were used (columns). The errors from top to bottom are for the lower limit of segmentation accuracy estimated in [10] (theoretical), labeled volumes (NMI based optimization), modified volumes (NMI) and real MR volumes using both NMI and Mahalanobis based optimization. ***Bottom:*** Comparison of optimization algorithms. Segmentation error and standard deviation (mm) was computed using PCA with 100 artificial modes.

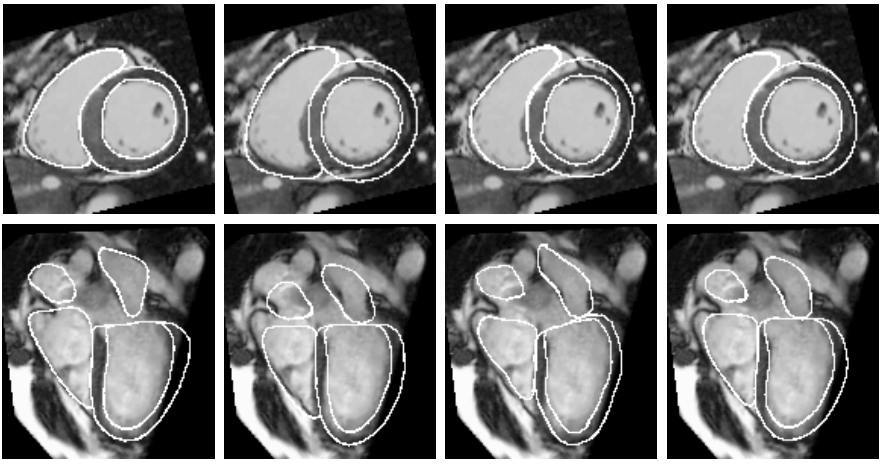
Method	Standard PCA	Artificial 24	Artificial 100	Artificial 250
Theoretical	-	1.50	1.03	0.70
Labeled	1.45 ± 0.48	1.42 ± 0.40	1.02 ± 0.36	1.00 ± 0.39
Modified	1.86 ± 0.48	1.67 ± 0.39	1.36 ± 0.30	1.33 ± 0.31
NMI	2.47 ± 0.52	2.54 ± 0.57	2.40 ± 0.43	2.28 ± 0.41
Mahalanobis	2.23 ± 0.44	2.26 ± 0.50	2.16 ± 0.41	2.11 ± 0.45

Algorithm	Labeled	Modified	NMI	Mahalanobis
Conjugate gradient	1.17 ± 0.37	1.43 ± 0.31	2.45 ± 0.49	2.21 ± 0.41
Simplex	1.03 ± 0.35	1.38 ± 0.29	2.39 ± 0.37	2.17 ± 0.43
Simulated annealing	1.02 ± 0.36	1.36 ± 0.30	2.40 ± 0.43	2.16 ± 0.41

was about 1.4 mm. This value is in accordance with the estimates obtained in [3] (1.8 mm) and in [4] (1.2 mm). The accuracy of automatic segmentation (Table 2) was about 1.9 mm for the ventricles and epicardium. In other words, the accuracy should be improved by 0.5 mm in order to reach the human accuracy. However, according to our results increasing the number of deformation modes above 100-200 does not solve the problem but the optimization of the weights of the deformation modes becomes the problem. Although global optimization was used, the accuracy could not be significantly improved. Two reasons can explain the result: 1) the function to be optimized does not have the global

Table 2. Segmentation error for different cardiac objects using simulated annealing and Mahalanobis based optimization criterion. The error and its standard deviation (mm) are obtained when using standard PCA (top) and PCA with 250 artificial modes (bottom). Abbreviations LV, RV, LA, RA and EC stand for left ventricle, right ventricle, left atrium, right atrium and epicardium, respectively. As shown in Table 1, total errors were 2.23 mm for standard PCA and 2.11 mm for PCA with 250 artificial modes.

	LV	RV	LA	RA	EC
St PCA	1.89 ± 0.36	2.23 ± 0.64	2.67 ± 0.76	3.21 ± 1.67	1.96 ± 0.40
Artif 250	1.77 ± 0.36	2.13 ± 0.63	2.44 ± 0.85	3.17 ± 1.89	1.85 ± 0.51



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

Fig. 2. Two randomly selected examples of the segmentations achieved by using Mahalanobis distance based criterion and simulated annealing. The slices with the short- (top) and long-axis (bottom) view have been chosen from different cases. a) Manual segmentations, b) initial conditions and c) automatic segmentations using standard PCA and d) PCA with 250 artificial modes. The errors for these particular cases were 1.65 and 1.61 (top) and 2.36 and 2.18 (bottom) for the standard PCA and PCA with 250 artificial modes, respectively.

maximum/minimum in the correct location due to image imperfections, and 2) the optimization was not run long enough to reach the global optimum. Despite the correct reason, increasing the computation time considerably is not feasible from clinical point of view. Therefore, more attention should be put on developing better optimization criteria. Furthermore, to ensure the artificial shapes produced by the non-rigid movement strategy (Eq. 2) would reflect true and physically plausible shape variation, it would be beneficial to constrain the deformations of the surface points using the local probability density functions generated from the database.

In terms of the energy optimization, simulated annealing outperformed simplex and conjugate gradient algorithms in all cases. However, since the difference in the segmentation errors was not statistically significant (Table 1), the choice between the optimization algorithm had no substantial role in the segmentation. The Mahalanobis distance performed slightly better than NMI when real MR volumes were used. This was probably due to small amount of information obtained from profile points.

For real MR volumes, the method developed resulted in relatively small final error, 2.11 mm by using 250 modes, simulated annealing and Mahalanobis-distance. The errors for the left and right ventricles were 1.79 mm and 2.14 mm, respectively. The errors are comparable with the results recently reported, and slightly better than reported for example in [3, 14].

References

1. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active shape models - their training and application. Computer Vision and Image Understanding **61** (1995) 38–59
2. Mitchell, S.C., Bosch, J.G., Lelieveldt, B.P.F., van der Geest, R.J., Reiber, J.H.C., Sonka, M.: 3-D active active appearance models: segmentation of cardiac MR and ultrasound images. IEEE Trans. Med. Imaging **21** (2002) 1167–1178
3. Lötjönen, J., Kivistö, S., Koikkalainen, J., Smutek, D., Lauerma, K.: Statistical shape model of atria, ventricles and epicardium from short- and long-axis MR images. Medical image analysis **8** (2004) 371–386
4. Assen, H., Danilouchkine, M.G., Frangi, A.F., Ordás, S., Westenberg, J.J.M., Reiber, J.H.C., Lelieveldt, B.P.F.: SPASM: A 3D-ASM for segmentation of sparse and arbitrarily oriented cardiac MRI data. Medical Image Analysis **10** (2006) 286–303
5. van Ginneken, B., Frangi, A., Staal, J., ter Haar Romeny, B., Viergever, M.: Active shape model segmentation with optimal features. IEEE Trans. Med. Imaging **21** (2002) 924–933
6. Lorenzo-Valdéz, 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** (2004) 255–265
7. Davatzikos, C., Tao, C., Shen, D.: Hierarchical active shape models using the wavelet transform. IEEE Trans. Med. Imaging **22** (2003) 414–423
8. Zhao, Z., Aylward, S.R., Teoh, E.K.: A novel 3D partitioned active shape model for segmentation of brain MRI images. In: Proc. MICCAI '05. (2005) 221–228
9. de Bruijne, M., van Ginneken, B., Viergever, M., Niessen, W.: Adapting active shape models for 3D segmentation of tubular structures in medical images. In: Proc. IPMI '03. (2003) 136–147
10. Lötjönen, J., Antila, K., Lamminmäki, E., Koikkalainen, J., Lilja, M.: Artificial enlargement of a training set for statistical shape models: Application to cardiac images. In: Proc. FIMH 2005. (2005) 92–101
11. Cootes, T.: Statistical modes of appearance for computer vision. Tech. report, Univ. of Manchester (2004)
12. Studholme, C., Hill, D., D.J.Hawkes: An overlap invariant entropy measure of 3D medical image alignment. Pattern Recognition **32** (1999) 71–86
13. Press, W., Teukolsky, S., Vetterling, W., Flannery, B.: Numerical recipes in C: the art of scientific computing. Second edn. Cambridge University Press (1992)
14. Kauss, M., von Berg, J., Weese, J., Niessen, W., Pekar, V.: Automated segmentation of the left ventricle in cardiac MRI. Medical Image Analysis **8** (2004) 245–254

Segmenting Lung Fields in Serial Chest Radiographs Using Both Population and Patient-Specific Shape Statistics

Yonghong Shi[1], Feihu Qi[1], Zhong Xue[2], Kyoko Ito[3], Hidenori Matsuo[3], and Dinggang Shen[2,*]

[1] Department of Computer Science and Engineering
Shanghai Jiao Tong University, Shanghai, China 200030
{shi-yh, fhqi}@cs.sjtu.edu.cn
[2] Section of Biomedical Image Analysis, Department of Radiology
University of Pennsylvania, Philadelphia, PA 19104,
{zhong.xue, dinggang.shen}@uphs.upenn.edu
[3] Hidaka Hospital, 886 Nakaomachi, Takasaki Gunma-Prefecture, 370-0001, Japan
{itokyon@da3.so-net.ne.jp, hidakadc}@cocoa.ocn.ne.jp

Abstract. This paper presents a new deformable model using both population-based and patient-specific shape statistics to segment lung fields from serial chest radiographs. *First*, a modified scale-invariant feature transform (SIFT) local descriptor is used to characterize the image features in the vicinity of each pixel, so that the deformable model deforms in a way that seeks for the region with similar SIFT local descriptors. *Second*, the deformable model is constrained by both population-based and patient-specified shape statistics. Initially, population-based shape statistics takes most of the rules when the number of serial images is small; gradually, patient-specific shape statistics takes more rules after a sufficient number of segmentation results on the same patient have been obtained. The proposed deformable model can adapt to the shape variability of different patients, and obtain more robust and accurate segmentation results.

1 Introduction

Evaluation of the variation of cardiac size from month to month by taking serial chest radiographs [1-3] remains crucial for the treatment of hemodialysis patients. This is because the evaluation of a patient's dry weight (DW) is usually based on the cardiac size measured during each dialysis session [4], and the misevaluation of DW often results in a fatal illness such as patient's death. Since segmentation of lung fields in the chest radiographs provides a means to measure cardiac size, it is important to accurately segment lung fields from the regularly captured serial chest radiographs of each patient. Methods for segmenting lung fields in chest radiographs [1] can be classified into three groups, i.e., the rule-based methods [3], the pixel-based classification methods [2], and the deformable model-based methods, including ASM [5] and AAM [6]. In this paper, a new deformable model is proposed for robust and accurate serial lung field segmentation.

* Corresponding author. E-mail: dinggang.shen@uphs.upenn.edu (Dinggang Shen).

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 83 – 91, 2006.

In deformable segmentation, it is important to characterize the relatively rich image features around the boundaries of lung fields, and then use these image features to guide the image segmentation. This is because the generic image features, such as edges along the boundaries of lung fields, are always inconsistent, and not sufficiently distinctive to discriminate, for example, rib edges from lung field edges. Therefore, the complex local descriptors, such as SIFT [7], might be suitable to characterize the image features around each point along the boundaries of lung fields. In the proposed deformable model, SIFT is used to capture image features for guiding the segmentation of lung fields, since it has been validated as the best among many local descriptors [8].

Another important idea in the proposed method is that the deformable model is constrained by both population-based and patient-specific shape statistics. The shape statistics collected from the segmentation results of a population, i.e., the shapes of lung fields of individuals, can be used for constraining the segmentation of lung fields in the initial time-point images of a specific patient. As the number of serial images of the same patient increases with time, it is possible to collect the patient-specific shape statistics from previous segmentation results of the same patient and use it to guide the segmentation of lung fields in the serial chest radiographs of the same patient. In this way, the deformable model gradually adapts to the shape statistics of the specific patient and yields more robust segmentation results. In our method, the patient-specific shape statistics is updated each time after a new image is acquired and segmented; then the segmentation results of all the previous time-point images are further refined by using the updated shape statistics.

2 Method

2.1 Description

In the clinical setting, serial chest radiographs are acquired monthly from each patient. Our goal is therefore to accurately segment the lung fields from the serial images and provide measures such as cardiac size for quantitative estimation of DW. At each time-point t, two major steps are performed for the newly captured image, i.e., the preprocessing step and the deformable segmentation step, as detailed next.

- The preprocessing step. The first time-point image is selected as the template image, and the image captured at time t is rigidly transformed onto the space of the template image by using rigid transformation [9]. Moreover, the image intensities are globally normalized to have a similar distribution to the one of the template image [10]. Thus, we obtain a normalized image at time t. All the normalized serial images are denoted as $\{I_i, i=1, 2,\ldots, t\}$, where I_t is the current time-point image.
- The deformable segmentation step. A new deformable model is utilized to segment the current image I_t, by using the statistical information collected from a population and also from t-1 previously segmented images of the same patient. The energy function of the deformable model is defined as,

$$E_t = E_t^{\mathbf{Smth}} + E_t^{\mathbf{SIFT}} + (1-\omega_t)E_t^{\mathbf{P}} + \omega_t E_t^{\mathbf{I}} \tag{1}$$

where $E_t^{\mathbf{Smth}}$ denotes the smoothness constraint of the deformable contour, and it requires that the first and the second derivatives of the deformable contour be close to

zero. $E_t^{\mathbf{SIFT}}$ denotes the overall difference of the SIFT features of the corresponding model points in the image I_t and in the template I_0, which is actually a mean shape generated from a public database [2, 11]. $E_t^{\mathbf{P}}$ denotes the constraint derived from the population-based shape statistics, while $E_t^{\mathbf{I}}$ denotes the constraint required by the patient-specific shape statistics, trained from t-1 segmentation results on previous time-point images of the same patient. The energy term $E_t^{\mathbf{I}}$ is not existent for several initial time-point images, simply because the available segmentation results for the specific patient are not sufficient to train a statistical shape model. We can actually begin to train the patient-specific statistical shape model once $N_{\mathbf{s}}$ time-point images have been collected and segmented. Here, $N_{\mathbf{s}}$ is the minimal number of segmentation results for the same patient, and $N_{\mathbf{s}} = 5$ is used in our study. Afterward, we can gradually increase the weight of energy term $E_t^{\mathbf{I}}$ and simultaneously decrease the weight of energy term $E_t^{\mathbf{P}}$, as more and more new images are added. This is controlled by parameter ω_t :

$$\omega_t = \begin{cases} 0 & t \le N_{\mathbf{s}} \\ (t - N_{\mathbf{s}})/(N_{\mathbf{b}} - N_{\mathbf{s}}) & N_{\mathbf{s}} < t \le N_{\mathbf{b}} \\ 1 & N_{\mathbf{b}} < t \end{cases} \tag{2}$$

If we have more than $N_{\mathbf{b}}$ images from the same patient, we will stop using the population-based shape constraint term $E_t^{\mathbf{P}}$, since the patient-specific statistics collected from over $N_{\mathbf{b}}$ samples is sufficient to capture the variation of lung fields of the patient.

The deformable segmentation of image I_t is summarized as follows:

(1) If $t \le N_{\mathbf{s}}$, segment I_t by minimizing Eq. (1), where ω_t is 0 and $E_t^{\mathbf{I}}$ is not used.

(2) If $t > N_{\mathbf{s}}$, the following two steps are performed (refer to Fig.1 for details):

Step (2.1), segment image I_t by incorporating the patient-specific constraint term $E_t^{\mathbf{I}}$, which is derived from t-1 segmentation results on previous time-point images $I_1,...,I_{t-1}$.

Step (2.2), update the patient-specific shape statistics by adding a new segmentation result of I_t to the training set, and then use this updated patient-specific shape statistics to refine the segmentations on all t images $I_1,...,I_t$. Finally, the patient-specific shape statistics is updated again, which will be used for segmenting the next time-point image.

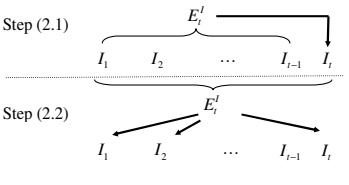


Fig. 1. Two major steps for segmenting the current image at t

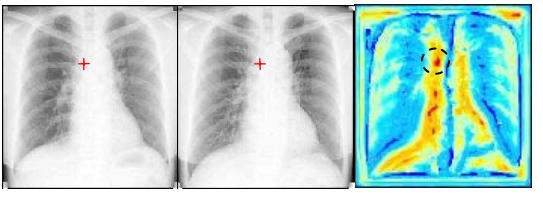


Fig. 2. Demonstration of using SIFT features for correspondence detection in two different time-point images

2.2 Energy Terms

Denoting the deformable contour as M ordered points, i.e., $\{v_i = (x_i, y_i) \mid i = 1,2,...,M\}$, the smoothness constraint term $E_t^{\mathbf{Smth}}$ can be represented as,

$$E_t^{\mathbf{Smth}} = \sum_{i=1}^{M} \left(\alpha_i E_{t,i}^{\mathbf{Cont}} + \beta_i E_{t,i}^{\mathbf{Curv}} \right) = \sum_{i=1}^{M} \alpha_i (| \bar{d}_t - \| v_{t,i} - v_{t,i-1} \| |) + \beta_i (\| v_{t,i-1} - 2v_{t,i} + v_{t,i+1} \|) \quad (3)$$

where $E_{t,i}^{\mathbf{Cont}}$ and $E_{t,i}^{\mathbf{Curv}}$ are the continuity and the curvature constraints on the i-th point of the contour during the segmentation of lung fields at time t. α_i and β_i are the weights, and $\bar{d}_t$ is the average distance between neighboring points in the contour.

The second energy term in Eq. (1) $E_t^{\mathbf{SIFT}}$ is a distance measure between two SIFT features, and minimizing it enables us to determine point correspondences between the template I_0 and image I_t. As we will describe in the next subsection, SIFT features are distinctive, thus facilitating the correspondence detection in the chest radiographs. The SIFT features used in our study are designed as: for each point v_i, a 128-element vector s_i is used to represent the histogram of orientations in the local image around the point v_i. Then, the energy term $E_t^{\mathbf{SIFT}}$ can be defined as,

$$E_t^{\mathbf{SIFT}} = \sum_{i=1}^{M} \prod_{l=1}^{128} \left| s_i^{\mathrm{Mdl}}(l) - s_i^t(l) \right| \quad (4)$$

where $s_i^{\mathbf{Mdl}}$ and s_i^t represent the SIFT feature vectors of point v_i in the template image I_0 and its corresponding point in the image I_t. $s_i^{\mathbf{Mdl}}(l)$ and $s_i^t(l)$ represent the l-th element of $s_i^{\mathbf{Mdl}}$ and s_i^t, respectively.

The third and forth energy terms are population-based and patient-specific statistical constraints, respectively. We use [5] to capture shape statistics from the population, i.e., a set of manually segmented training samples [2]. In contrast, we use a hierarchical shape representation in [12] to effectively capture the patient-specific shape statistics, since the number of segmentation results from the same patient is not large.

2.3 SIFT Local Descriptor

SIFT, as detailed in [7], consists of four major steps: (1) scale-space peak selection; (2) key point localization; (3) orientation assignment; (4) key point description. The first two steps detect the key points in the scale space, and they are not required in this study since we only focus on using SIFT features of the contour points along the boundaries of lung fields. Thus, we use the last two steps to compute the SIFT features for each contour point. The SIFT local descriptor for a point is computed as follows:

(1) A patch is centered on the point, rotated according to the dominant gradient orientation around the point, and scaled to an appropriate size.
(2) The SIFT local descriptor is then created by sampling the magnitudes and the orientations of the image gradients in the patch around the point, to build the smoothed orientation histograms for capturing the important information in this patch of image. A 4x4 array of histograms, each with 8 orientation bins, is obtained, capturing the rough spatial structure of this patch of image.

(3) This 128-element vector is then normalized to a unit length and further thresholded to remove the elements with small values.

Fig. 2 shows the performance of SIFT features in detecting corresponding points in the chest radiographs. In this figure, the SIFT features of the red crossed point in one time-point image as shown in the left image are compared with the SIFT features of all the points in another time-point image as shown in the middle image. As indicated by the color-coded similarity map in the right image, a very small number of points in the middle image are similar to the red crossed point in the left image. Since the two images have been rigidly aligned, the corresponding point of the red crossed point in left image is detected in the middle image by searching the maximal similarity within a local neighborhood.

2.4 Learning Patient-Specific Shape Statistics

For serial image segmentation, it is important to use the segmentation results from the previous time-point images to guide the segmentation of the current time-point image. Similarly, the segmentation results of previous images can also be refined using the information collected in the later time-points. Since we have normalized intensities and poses of other time-point images to the first time-point image, all the lung fields are globally overlapped well. However, although the globally aligned serial lung field shapes are very similar, there still exist visible differences, due to distortion, pose change, 3D-2D projection, and normal or abnormal development. Fig. 3 shows an example of the serial lung field shapes. The thin black contours represent the boundaries of lung fields in each image, while the thick red contours are identical and they are actually the same contour segmented from the first time-point image. It can be seen that the segmentation results obtained from the serial images of the same patient are globally similar while locally different, because of the reasons mentioned above. Note that it is relatively easy to capture the shape statistics from the shapes of the same patient, even using a small number of shape samples. This type of patient-specific shape statistics, collected from the previous segmentation results, can guide the segmentation of the current and future time-point images. This is the main idea of our on-line learning method for capturing the patient-specific shape statistics.

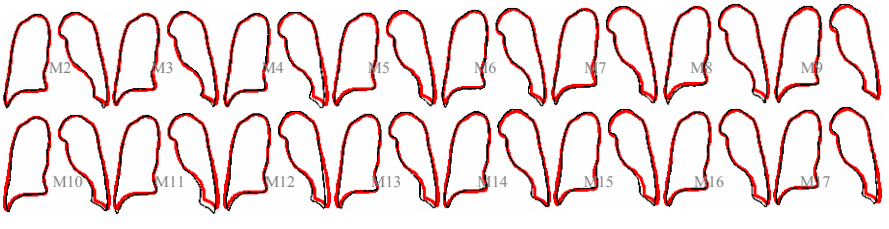


Fig. 3. Demonstration of local deformations of lung field shapes from the serial images of the same patient. The thin black contours are the segmentation results of the serial images, and the thick red contours are identical, denoting the same shape segmented from the first time-point image.

At time-point t, we have t-1 previously segmented images and a current time-point image I_t. We can use the t-1 segmentation results from images $I_1,...,I_{t-1}$ (in Step (2.1) of Fig.1), or the t segmentation results from images $I_1,...,I_t$ (in Step (2.2) of Fig.1), as the training samples. In order to capture the patient-specific shape variability information from the training samples, we adopt the hierarchical representation of shape statistics that can effectively guide the deformable segmentation of lung fields in the current time-point image [12].

It should be noted that this on-line and incrementally learned patient-specific shape statistics can be used only after a sufficient number (N_s) of segmentation results are obtained. As more images are processed, the constraint derived from the population-based shape statistics becomes less and less important, while the constraint derived from the patient-specific shape statistics increases gradually. Finally, the weight for population-based statistical constraint is 0 if we have N_b or more training samples. This mechanism enables the deformable model to adapt to the patient-specific shape variability, thus it is more robust and accurate for lung field segmentation.

3 Experimental Results

The serial frontal chest radiographs of 39 patients, most with up to 17 monthly scans, are used. The lung fields of all serial images have been manually delineated by a human observer, which are used as a gold standard to validate the segmentation results obtained by our method, a standard Snake [13], and ASM [6] in the next.

3.1 Segmentation of Serial Chest Radiographs

Visual comparison. Fig.4 qualitatively compares the performances of our method, the standard Snake, and ASM in segmenting the lung fields from the serial chest radiographs (256x256). Initializations for these three methods are displayed in the top left image. The top right image is a selected time-point image from the serial images. The red, green and black contours denote the segmentation results by the standard Snake, ASM, and our method, respectively. It can be observed that our results are very close to the lung field boundaries, and they are temporally more stable than the ASM results (please refer to the four small serial images shown on the bottom of the figure). These results indicate that the use of patient-specific shape statistics improves the accuracy of image segmentation. They also indicate that a simple feature, such as a local gradient used in the standard Snake, is not able to describe the complex and inconsistent image content along the boundaries of lung fields. On the other hand, SIFT features facilitate the detection of correspondences in the serial images, which ensures the relatively better segmentations by our method.

Quantitative comparison. We also used quantitative measures to compare the segmentation results obtained by the three segmentation algorithms. We use the overlay percentage to compare the automated segmentation results with the manual segmentation results for each image. It turns out that the mean and standard deviation of overlay percentage are 93.8%±0.58% by our method, 92.6%±0.8% by ASM, and 78.7%±1.3% by the standard Snake. We also compute the average contour distance between automated segmentations and manual segmentations [2]. The mean and standard deviation

of average contour distance are 1.76±0.23 pixels by our method, 2.06±0.25 pixels by ASM, and 5.31±0.29 pixels by the standard Snake, respectively. These results show that our method achieved better performance than the standard Snake and ASM.

3.2 1D and 2D Cardiothoracic Ratios

By segmenting lung fields from chest radiographs, we can compute the diagnostic measures, i.e., cardiothoracic ratio (CTR), referred to as 1D-CTR. 1D-CTR is defined as the ratio of the transverse diameter of the heart to the transverse diameter of the thorax, as shown in Fig.5. Recently, the extraction of 2D cardiothoracic ratio (2D-CTR) from the chest radiographs has been extensively studied [14], since it is believed to be more robust than 1D-CTR. In our study, 2D-CTR is defined as the squared root of the ratio between the area of white region and the area of both white and grey regions, as shown on the right of Fig.5.

The importance of 1D-CTR has been validated in many clinical applications. For evaluating the usefulness of 2D-CTR, we need to measure the relationship between 1D-CTR and 2D-CTR from the same serial chest radiographs. As shown by the left plot in Fig.6, 2D-CTR is highly related to 1D-CTR, which indirectly indicates that 2D-CTR can be also used as a diagnostic measure.

Moreover, the Bland and Altman plots [15] of 1D-CTR and 2D-CTR are provided in the middle and right plots in Fig.6, respectively, by using the manual segmentation results as the gold standard. These two plots indicate that the distribution of 2D-CTR is more compact than that of 1D-CTR. Also, the mean difference value by 2D-CTR measurement is smaller than that by 1D-CTR measurement. Moreover, the confidence interval of 2D-CTR is tighter than that of 1D-CTR.

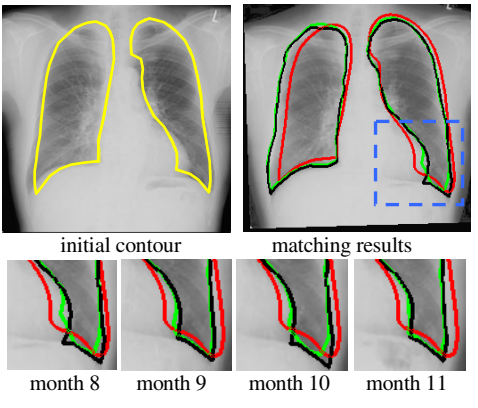


Fig. 4. Comparison of automated segmentations

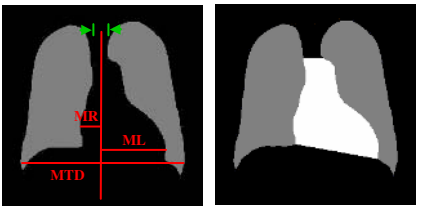


Fig. 5. 1D-CTR and 2D-CTR. In left image, MR+ML denotes the cardiac diameter, and MTD denotes the thoracic diameter. In right image, white region represents the central shadow in the chest radiograph, while gray region represents the bilateral lung fields.

4 Conclusion

We have presented a new deformable model to segment the lung fields from serial chest radiographs by using both population-based and patient-specific shape statistical constraints. The patient-specific shape statistics is on-line and incrementally learned

from the segmentation results of the previous time-point images of the same patient. For the initial time-point images, the population-based shape statistics plays the major rule for statistically constraining the deformable contours. As more and more segmentation results are obtained, the patient-specific shape statistics starts to constrain the segmentation, and it gradually takes the major rule for statistical constraining. Also, SIFT features, used to characterize image points, are relatively distinctive, thus making the correspondence detection in the two chest radiographs relatively easy. Experimental results show that our method can segment the lung fields more accurately, compared to the standard Snake and ASM. Finally, we demonstrated that 2D-CTR is highly related to 1D-CTR, but it is more robust to compute the diagnostic measure.

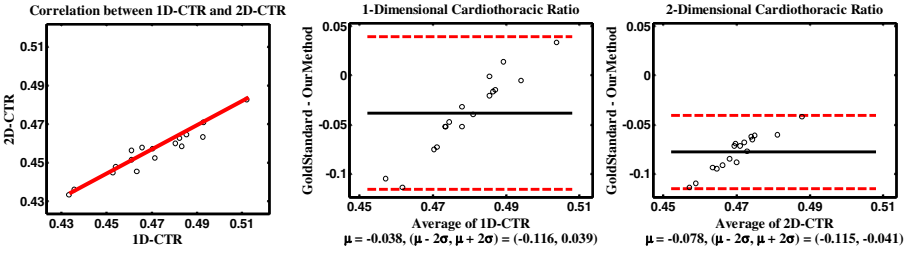


Fig. 6. Relationship between 1D-CTR and 2D-CTR (left), the Bland and Altman plots of 1D-CTR and 2D-CTR (middle and right), respectively

References

1. B. van Ginneken, B.M. ter Haar Romeny, M.A. Viergever. Computer-aided diagnosis in chest radiography: a survey. *IEEE Trans. on Med Imaging*, 20(12): 1228-1241, 2001.
2. B. van Ginneken, M.B. Stegmann, and M. Loog. Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. *Medical Image Analysis*, 10:19-40, 2006.
3. M.S. Brown, L.S. Wilson, B.D. Doust, R.W. Gill, and C. Sun. Knowledge-based method for segmentation and analysis of lung boundaries in chest X-ray images. *Computerized Medical Imaging and Graphics*, 22:463-477, 1998.
4. J.W. Jaeger and R.L. Mehta. Assessment of dry weight in hemodialysis: an overview. *Journal of American Social Nephrology*, 10:392-403, 1999.
5. T.F. Cootes, C.J. Taylor, D. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38-59, 1995.
6. T.F. Cootes and C.J. Taylor. Statistical models of appearance for computer vision. *Technical Report, Wolfson Image Analysis Unit, University of Manchester*, 2001.
7. D.G. Lowe. Object recognition from local scale-invariant features. In *International Conference on Computer Vision*, 1999.
8. K. Mikolajczyk and C.Schmid. A performance evaluation of local descriptors. In *IEEE conference on Computer Vision and Pattern Recognition*, 2003.
9. M. Jenkinson and S. Smith. A global optimization method for robust affine registration of brain images. *Medical Image Analysis*, 5:143-156, 2001.
10. L.G. Nyul, J.K. Udupa, and X. Zhang. New variants of a method of MRI scale standardization. *IEEE Trans. on Medical Imaging*, 19(2): 143-150, 2000.

11. Shiraishi, J., Katsuragawa, S., Ikezoe, J., Matsumoto, T., Kobayashi, T., Komatsu, K., matsui, M., fujita, H., Kodera, Y., Doi, K., Development of a digital image database for chest radiographs with and without a lung nodule: receiver operation characteristic analysis of radiologists' detection of pulmonary nodules. *American Journal of Roentgenology*, 174: 71-74, 2000.
12. C. Davatzikos, X. Tao, and D. Shen. Hierarchical active shape models using the wavelet transform. *IEEE transaction on medical imaging*, 22(3):414-423, 2003.
13. M. Kass, A.Witkin, and D. Terzopoulos. Snakes: active contour models. *Internal Journal of Computer Vision*, 1(4): 321-331, 1987.
14. R.F.J. Browne, G. O'Reilly, and D. Mclnerney. Extraction of the two-dimensional cardiothoracic ratio from digital PA chest radiographs: correlation with cardiac function and the traditional cardiothoracic ratio. *Journal of Digital Imaging*, 17(2): 120-123, 2004.
15. J.M. Bland and D.G. Altman. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet* 346 (8982), 1085-1087, 1995.

4D Shape Priors for a Level Set Segmentation of the Left Myocardium in SPECT Sequences

Timo Kohlberger[1], Daniel Cremers[2], Mikaël Rousson[1], Ramamani Ramaraj[1], and Gareth Funka-Lea[1]

[1] Siemens Corporate Research, Inc., Imaging and Visualization Department, Princeton, NJ, USA
Timo.Kohlberger@siemens.com
[2] Department of Computer Science, University of Bonn, Germany
dcremers@cs.uni-bonn.de

Abstract. We develop a 4D (3D plus time) statistical shape model for implicit level set based shape representations. To this end, we represent hand segmented training sequences of the left ventricle by respective 4-dimensional embedding functions and approximate these by a principal component analysis. In contrast to recent 4D models on explicit shape representations, the implicit shape model developed in this work does not require the computation of point correspondences which is known to be quite challenging, especially in higher dimensions. Experimental results on the segmentation of SPECT sequences of the left myocardium confirm that the 4D shape model outperforms respective 3D models, because it takes into account a statistical model of the temporal shape evolution.

1 Introduction

Model-based imaging analysis of the left ventricle (LV) has gained an important role in diagnosis and treatment of heart diseases [8]. Segmentation in particular, either of the inner volume or the myocardium has found to be a significant prerequisite of further quantitative analysis, such as the estimation of the ejection fraction within one cardiac cycle. With standard 3D approaches, segmentation is carried out for each recorded image volume separately, i.e. by neglecting all temporal correlations. Recent approaches however aim at taking also the temporal dimension into account, arising from the fact that more and more 4D, i.e. 3D plus time, imaging techniques have become available. Relative to cardiac imaging, which provides a relatively high three-dimensional spatial resolution, isotope imaging methods, like gated perfusion SPECT, still have low resolution in both space and time and exhibit a relatively low signal to noise ratio. On the other hand, SPECT imaging inherently provides an excellent mean for cardiovascular diagnosis, because it allows direct visualization of physiology.

In terms of shape modeling, either explicit representations by landmarks or binary images/volumes are employed, or implicit ones, describing the separating three-dimensional contour as the zero-level set of a four dimensional function [16, 11]. Additionally, a statistical model is incorporated, in order to describe

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 92–100, 2006.

inter-patient shape variabilities. In most cases, Principal Component Analysis (PCA) is the method of choice. In the case of explicit shape representations [4, 9], PCA is either applied directly to the landmark coordinates or to the components of a deformation field relative to a mean shape, or, for the implicit representation, to the components of the embedding level set functions [11, 17, 20].

The extension of shape models to the spatio-temporal case has been mainly pursued with explicit (parametric) shape models [18, 3, 12, 14]. While these methods have shown quite some success in modeling and estimating the temporal evolution of surfaces, one should point out that explicit shape representations are known to have certain inherent limitations. Firstly, when matching two parametric surfaces (in the shape learning phase), there is the fundamental correspondence issue of which points on one surface are associated with which points on the other surface. Explicit representations are typically based on a specific point correspondence, estimating this correspondence and thus solving this combinatorial problem is an enormous computational challenge [6], especially in the 4D case considered in this paper. Misaligned point correspondences will appear as artifacts in the subsequently estimated shape deformation modes. Secondly, in the shape estimation step, the evolution of parametric surfaces has certain limitations. The spatial resolution of the surface representation is determined by the local density of control points which may fluctuate due to the motion of points. This evolution may become unstable once control or marker points start overlapping. Splitting or remerging of contours in the segmentation of multiply-connected structures is by default not possible. Sophisticated regridding mechanisms are required to handle these problems [7]. At the same time, the required amendments to the control point evolution may lead to artifacts in the shape deformation that are not inherent to the geometric shape warping problem.

While the level set method overcomes the above drawbacks of explicit representations, there have only been very few recently published works on extending the shape modeling to the spatio-temporal domain [5, 19]. While the latter approaches treat the temporal dimension separately and employ dynamical models, we propose to treat time as an ordinary fourth dimension and thus implementing a segmentation of the volume sequence as a whole. Unlike Chandrashekara et al. [3], who developed this idea for deformation fields relative to an explicit (parametric) mean sequence, we propose to apply a 4D PCA subspace estimation directly to the level set functions. As a consequence, we do not require a sophisticated non-rigid registration of training sequences, or the challenging combinatorial correspondence searches required by explicit landmark-based approaches. Moreover, due to considering all volumes of a sequence at the same time, and thereby also learning the temporal evolution of the shape model, segmentation quality can be shown to improve over treating each volume separately [11]. In addition, temporal alignment can be incorporated into the segmentation process straightforwardly, avoiding the need for any pre-alignment of the input sequence.

2 Approach

2.1 Level Set Representation

Let the recorded image volume sequence be denoted by the continuous intensity function $I : \Omega \times [1, T] \rightarrow \mathbb{R}$ defined on the 3D volume $\Omega \in R^3$ and the time period $[1, T]$. In contrast to other approaches treating intensity volumes at different times points separately, in the following we will always consider I as a function in the four-dimensional spatio-temporal space $\Omega \times [1, T]$.

In order to be independent of contour parametrization, support topological changes as well as avoiding correspondence problems when introducing a prior shape model later on, we pursue the approach in [16] in application to the four-dimensional space and embed any contour $C : [0, 1]^2 \times [1, T] \Rightarrow \mathbb{R}^2$ as the zero level set of a function $\phi : \Omega \times [1, T] \Rightarrow \mathbb{R}$, such that

$$C(\mathbf{x}, t) = \{(\mathbf{x}, t) \mid \phi(\mathbf{x}, t) = 0\} \ ,$$

which has been successfully applied to 2D [13] and 3D [10] image segmentation. In aiming at having a unique correspondence between a shape C and its embedding function, we furthermore assume ϕ to be a signed distance function, i.e. $|\nabla \phi| = 1$ almost everywhere.

However, since we want to estimate spatial translation and scale as well as temporal translation and scale separately, we introduce the intermediate transformation

$$T_\theta(\mathbf{x}, t) := \begin{pmatrix} \theta_{m_s} \mathbf{x} - \boldsymbol{\theta}_{d_s} \\ \theta_{m_t} t - \theta_{d_t} \end{pmatrix} \ ,$$

with $\theta_{d_s} \in \mathbb{R}^3$ and $\theta_{d_t} \in \mathbb{R}$ referring to translation parameters in space and time, respectively, and $\theta_{m_s}^{-1} \in \mathbb{R}$ and $\theta_{m_t}^{-1} \in \mathbb{R}$ to those in scale[1], respectively. By this, the embedding function will, in the following, be of the form:

$$\phi_\theta(\mathbf{x}, t) := \theta_{m_s}^{-1} \ \phi\big(T_\theta(\mathbf{x}, t)\big) \ .$$

2.2 Segmentation as Variational Problem

Subsequently, the problem of segmenting the image sequence I implicitly by the 4D level set function ϕ is stated as the joined minimization of the energies

$$E(\phi, \theta) = E_{data}(\phi, \theta) + \nu \ E_{shape}(\phi) \ , \tag{1}$$

where the data-driven energy term $E_{data}(\phi, \theta) = -\log \mathcal{P}(I \,|\, \phi, \theta)$ measures the probability of observing the image sequence I given the segmentation ϕ (at place θ_d and scale θ_m), and $E_{shape}(\phi) = -\log \mathcal{P}(\phi)$, which reflects the probability of the occurrence of a segmentation ϕ in relation to a set of 4D training segmentations $\{\hat{\phi}_1, \ldots, \hat{\phi}_M\}$.

[1] Here, we consider only isotropic scale in space.

In particular, for the data energy term we make use of the global approach by Mumford and Shah [15] given as a level set formulation [2] in the spatio-temporal space $\Omega \times [0,T]$:

$$E_{data}(\phi,\theta) := \int_{[1,T]}\int_{\Omega} \Big(I(T_\theta(\mathbf{x},t)) - \mu_1(t)\Big)^2 \, H\phi_\theta(\mathbf{x},t) + \Big(I(T_\theta(\mathbf{x},t)) - \mu_2(t)\Big)^2 \, \Big(1 - H\phi_\theta(\mathbf{x},t)\Big) \, d\mathbf{x}\, dt \quad (2)$$

with $H\phi := H(\phi)$ denoting the Heaviside step function, which is one at coordinates $(\mathbf{x},t)$ where ϕ is positive, and zero else, as well as $\mu_1 = \frac{1}{\alpha_1}\int_{[1,T]}\int_{\Omega} IH\phi_\theta \, d\mathbf{x}\, dt$, $\mu_2 = \frac{1}{\alpha_2}\int_{[1,T]}\int_{\Omega} I(1-H\phi_\theta)\, d\mathbf{x}\, dt$, and $a_k = \int_{[1,T]}\int_{\Omega} H\phi_\theta \, d\mathbf{x}\, dt$, $k = 1,2$.

With respect to the shape prior, we propose to conduct an PCA directly on the set of vectorized training *sequences* thereby reaching a low-dimensional parametrization of the subspace [11, 20]:

$$\phi_{\boldsymbol{\alpha}}(\mathbf{x},t) := \phi_0(\mathbf{x},t) + \boldsymbol{\alpha}^\top \boldsymbol{\psi}(\mathbf{x},t) \quad (3)$$

where $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_K)^\top$ denotes the eigenmode vector of the 4D levelset functions and $\phi_0 = \frac{1}{M}\sum_{i=1}^{M} \hat{\phi}_i$ the mean function[2]. Note that in contrast to [20], the eigenmodes here describe *spatio-temporal* variations across the whole volume sequence. Within this subspace coordinate system, we assume the prior probability $\mathcal{P}(\phi,\theta)$ as multivariate Gaussian distribution, and infer for an arbitrary function $\phi_{\boldsymbol{\alpha}}$ the energy

$$E_{shape}(\phi) := \boldsymbol{\alpha}^\top \Sigma^{-1} \boldsymbol{\alpha}$$

with Σ^{-1} denoting the inverse eigenvalue matrix resulting from the PCA.

Moreover, following the approach suggested in [20], we represent ϕ by means of the parametrization given in (3) throughout the *whole* segmentation problem, i.e. also in the data term, and consequently substitute ϕ by $\phi_{\boldsymbol{\alpha}}$ in (2). Thereby, we restrict the minimization of (1) to the much lower-dimensional subspace induced by the training shapes.

Due to the prior shape model being invariant towards translation and scale both in space and time, it is required for the training set to be free of such transformations prior to applying the PCA. Here, this is implemented by a prior alignment of each training shape $\hat{\phi}_i$ to an arbitrarily chosen reference by minimizing the energy

$$\int_{[1,T]}\int_{\Omega} \Big(\theta_{m_s}^{-1}\phi_{ref}\big(T_\theta(\mathbf{x},t)\big) - \hat{\phi}_i(\mathbf{x},t)\Big)^2 d\mathbf{x}\, dt, \quad i = 1,\ldots,M\,.$$

2.3 Gradient Descent Minimization

Determining a minimum of (1) with respect to $\boldsymbol{\alpha}$ and θ is implemented by gradient descent with respect to both vectors, that is

[2] Note that $\phi_{\boldsymbol{\alpha}}$ is not a distance function, since the space of the latter is nonlinear.

$$\frac{d\boldsymbol{\alpha}(\tau)}{d\tau} = -\frac{\partial E_{data}(\boldsymbol{\alpha},\theta)}{\partial \boldsymbol{\alpha}} - \nu \frac{\partial E_{shape}}{\partial \boldsymbol{\alpha}} \;, \quad \text{and} \quad \frac{d\theta(\tau)}{d\tau} = -\frac{\partial E_{data}(\boldsymbol{\alpha},\theta)}{\partial \theta}$$

with τ denoting the artificial evolution time, as opposed to the physical time t, and $d\tau$ the gradient step size. Thereby, the data term gradient reads

$$\frac{\partial E_{data}}{\partial \alpha_i} = \int_{[1,T]} \int_{\Omega} \Big(\big(I-\mu_1\big)^2 - \big(I-\mu_2\big)^2 \Big)\; \delta(\phi_{\boldsymbol{\alpha},\theta})\; \theta_{m_s}^{-1} \psi_i\big(T_\theta(\mathbf{x},t)\big)\; d\mathbf{x}\; dt \;,$$

with δ referring to the regularized Dirac measure, and the prior gradient $\frac{\partial E_{shape}}{\partial \boldsymbol{\alpha}} = \Sigma^{-1}\boldsymbol{\alpha}$. Furthermore, the gradient of the scaling and translating transformation is given by

$$\frac{\partial E_{shape}}{\partial \theta_\xi} = \int_{[1,T]} \int_{\Omega} \big((I-\mu_1)^2 - (I-\mu_2)^2 \big)\, \delta(\phi_{\boldsymbol{\alpha},\theta})\; \nabla \phi_{\boldsymbol{\alpha},\theta}\; \frac{dT_\theta}{d\theta_\xi}\; d\mathbf{x}\; dt$$

with $\xi \in \{d, m\}$, and $\frac{dT_\theta}{d\theta_d} = -\begin{pmatrix} \mathbf{1}_3 \\ 1 \end{pmatrix}$, $\frac{dT_\theta}{d\theta_m} = \begin{pmatrix} \mathbf{x} \\ t \end{pmatrix}$, respectively.

3 Experimental Evaluation and Discussion

We compared the segmentation results of our proposed approach to that of Tsai et. al. [20] (3D, similarity transforms only, Chan-Vese) while measuring the *relative symmetric voxel error*: $|b(\phi) \otimes b(\phi_{ref})| \,/\, |b(\phi_{ref})|$ with $b(\phi)$ giving the binarized volume of ϕ, $\otimes$ denoting the exclusive or operator and ϕ_{ref} the ground truth distance function. As training and ground truth data sets we had 15 hand-segmented gated perfusion SPECT recordings of the left myocardium with a volume of 30×30×27 voxels (each of size 6.59 mm^3) times eight frames over the cardiac cycle, and possible pathologies being removed. After alignment due to translation and scale, as explained above, 3D PCA was conducted over the vectorized 15× 8 volumes alike, whereas 4D PCA on the 15 vectorized sequences only. Ten gradient descent iterations were taken for 3D or 4D. Scaling and translation parameters θ were performed prior to the interleaved descent for both $\boldsymbol{\alpha}$ and θ in order to avoid local minima with respect to $(\boldsymbol{\alpha}, \theta)$ due to spatial or temporal misalignments. Different weightings ν of the prior gradient were tried in the range from 1/100 to 1/1 of the data gradient's L^2 norm and we chose the one yielding the best results w.r.t. the quality measure.

Figures 1 and 2 depict two typical segmentation results for both methods in comparison to the hand-segmented ground truth. While the dataset in Fig. 1 was contained in the training set, that in Fig. 2 was not. In the first experiment, the 3D-based method yields a systematic over-segmenting especially at end diastole, which is mainly due to over-estimating the per-volume scales. The error for the 3D approach did range from 35% (end systole) up to 90% (end diastole) vs. 40% to 20% with the 4D approach (averages: 53% vs. 28%). In the second experiment, Fig. 2, the 3D prior subspace allows for too many variations[3], due to ignoring

[3] A stronger weighting of the prior term inhibits most of the temporal variation.

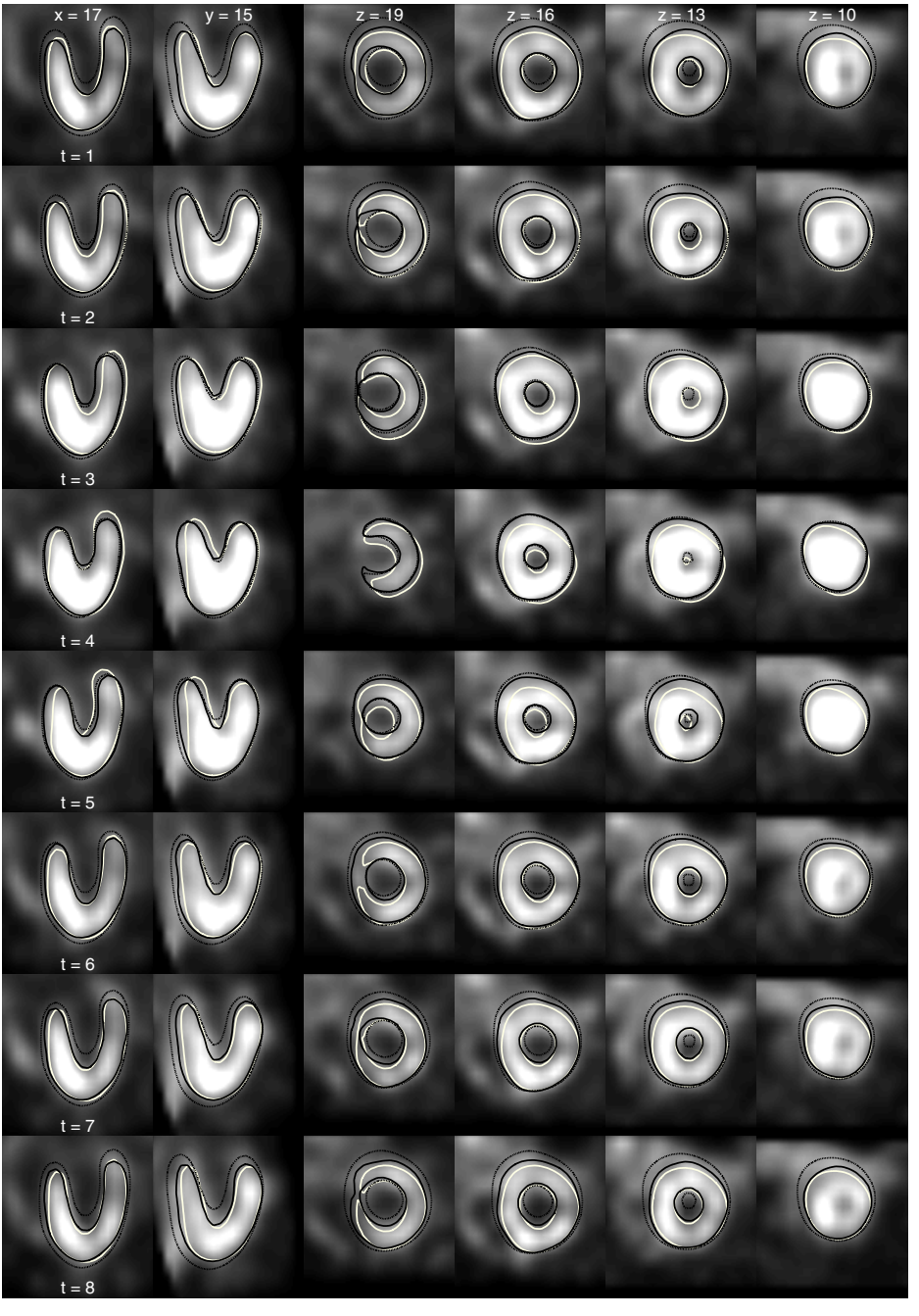


Fig. 1. Segmentation result for a sequence included in the training set. White: Manually segmented ground truth. Gray: 3D-prior method. Black: 4D-prior method. Estimating one scale for the whole sequence yields better results than one for each volume, and in turn allows for a better fit with respect to shape.

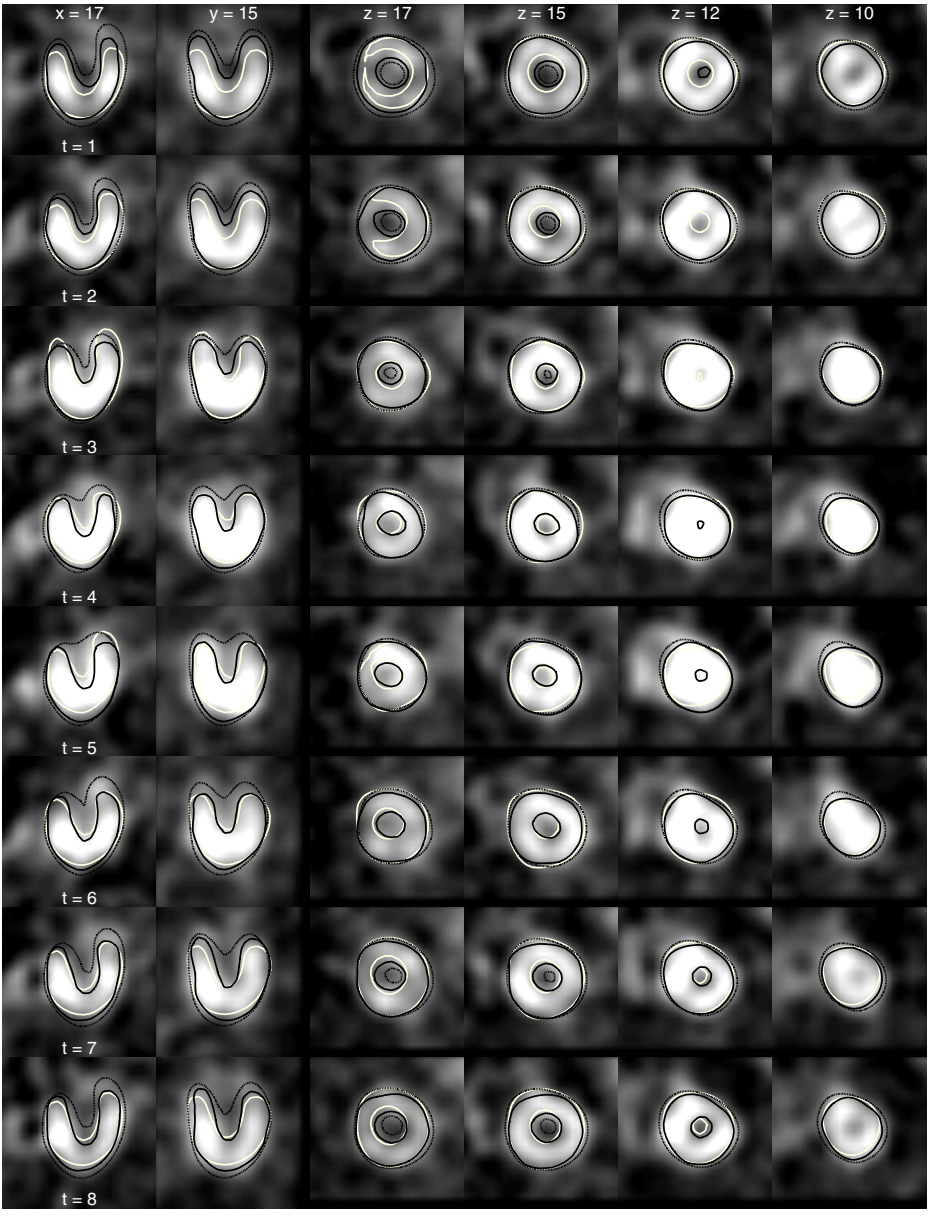


Fig. 2. Segmentation result for a sequence not being in the training set. White: Manually segmented ground truth. Gray: 3D-prior method. Black: 4D-prior method. Due to taking the temporal correlations into account also, the 4D prior yields a more specific prior per time point. Despite a strong weighting of the 3D prior, better results for the latter were not achievable, while still allowing for temporal variations.

the temporal index. On the other hand, the fifteen training vectors of the 4D PCA already yield an acceptable generalization (the errors were: 35% to 80% vs. 40% to 30% and 57% and 27% in average). These and other experiments clearly show the advantage of the new approach of taking the whole temporal dimension into account — both in the data-dependent probability, as well as the statistical shape prior.

Acknowledgements. The authors would like to thank Günther Platsch and Darrell Burckhardt of Siemens Molecular Imaging.

Upon submission of the camera-ready version of this manuscript, the authors of [1] have pointed us to a recently submitted related approach.

References

1. A. Abufadel, T. Yezzi, and R. Schafer. 4D segmentation of cardiac data using active surfaces with spatiotemporal shape priors. *Applied Pattern Recognition*. submitted.
2. T.F. Chan and L.A. Vese. Active contours without edges. *IEEE Trans. Image Processing*, 10(2):266–277, 2001.
3. R. Chandrashekara, A. Rao, G.I. Sanchez-Ortiz, and D. Rueckert. Construction of a statistical model for cardiac motion analysis using nonrigid image registration. In *18th Int. Conf. on Information Processing in Medical Imaging (IPMI '03)*, 2003.
4. T.F. Cootes, C. Beeston, G.J. Edwards, and C.J. Taylor. A unified framework for atlas matching using active appearance models. In *16th Int. Conf. Information Processing in Medical Imaging*, volume 1613, pages 322–333. Springer Berlin, 1999.
5. D. Cremers and G. Funka-Lea. Dynamical statistical shape priors for level set based sequence segmentation. In *Third Workshop on Variational and Level Set Methods*, volume 3752 of *LNCS*, pages 210–221. Springer Berlin, 2005.
6. R. Davies, C. Twining, Cootes T., J. Waterton, and C. Taylor. A minimum description length approach to statistical shape modelling. *IEEE Trans. on Medical Imaging*, 21(5):525–537, 2002.
7. H. Delingette and J. Montagnat. New algorithms for controlling active contours shape and topology. In *ECCV 2000*, volume LNCS 1843, pages 381–395. Springer.
8. A.F. Frangi. Three-dimensional modeling for functional analysis of cardiac images: A review. *IEEE Trans. Medical Imaging*, 20(1):2–25, Jan. 2001.
9. A.F. Frangi, D. Rueckert, J.A. Schnabel, and W.J. Niessen. Automatic construction of multiple-object three-dimensional statistical shape models: Applications to cardiac modeling. *IEEE Trans. Medical Imaging*, 21(9):1151–1166, Sept. 2002.
10. S. Kichenassamy, A. Kumar, P.J. Olver, A. Tannenbaum, and A.J. Yezzi. Gradient flows and geometric active contour models. In *ICCV 1995*, pages 810–815, 1995.
11. M.E. Leventon, W.E. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *CVPR 2000*, volume 1, pages 1316–1323. IEEE, 2000.
12. M. Lorenzo-Valdéz, G.I. Sanchez-Ortis, R. Mohiaddin, and D. Rueckert. Segmentation of 4d cardiac MR images using a probabilistic atlas and the EM algorithm. In *MICCAI 2005*, volume 2878, pages 440–450. Springer, Berlin, 2005.
13. R. Malladi, J. A. Sethian, and B.C. Vemuri. Shape modeling with front propagation: A level set approach. *PAMI*, 17(2):158–175, 1995.
14. J. Montagnat and H. Delingette. 4D deformable models with temporal constraints: Application to 4D cardiac image segmentation. *Med. Im. Anal.*, 9(1):87–100, 2005.

15. D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–685, 1989.
16. S.J. Osher and J.A. Sethian. Fronts propagation with curvature dependent speed: Algorithms based on hamilton-jacobi formulations. *J.of Comp. Ph.*, 79:12–49, 1988.
17. N. Paragios. A variational approach for the segmentation of the left ventricle in cardiac image analysis. *IJCV*, 50(3):345–362, July 2002.
18. D. Perperidis, R. Mohiaddin, and D. Rueckert. Construction of a 4d statistical atlas of the cardiac anatomy and its use in classification. In J. Duncan and G. Gerig, editors, *MICCAI 2005*, volume 3750 of *LNCS*, pages 402–410. Springer, 2005.
19. W. Sun, M. Cetin, R. Chan, V. Reddy, G. Holmvang, V. Chandar, and A. Willsky. Segmenting and tracking the left ventricle by learning the dynamics in cardiac images. Technical Report LIDS 2642, MIT, Feb. 2005.
20. A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, W.E. Grimson, and A. Willsky. A shape-based approach to the segmentation of medical imagery using level sets. *IEEE Trans. Medical Imaging*, 22(2):137–154, Feb. 2003.

Cell Segmentation Using Coupled Level Sets and Graph-Vertex Coloring

Sumit K. Nath, Kannappan Palaniappan, and Filiz Bunyak

MCVL, Department of Computer Science, University of Missouri-Columbia,
Columbia MO 65211, USA*
{naths, palaniappank, bunyak}@missouri.edu

Abstract. Current level-set based approaches for segmenting a large number of objects are computationally expensive since they require a unique level set per object (the N-level set paradigm), or $\lceil \log_2 N \rceil$ level sets when using a multiphase interface tracking formulation. Incorporating energy-based coupling constraints to control the topological interactions between level sets further increases the computational cost to $O(N^2)$. We propose a new approach, with dramatic computational savings, that requires only four, or fewer, level sets for an arbitrary number of similar objects (like cells) using the Delaunay graph to capture spatial relationships. Even more significantly, the coupling constraints (energy-based and topological) are incorporated using just constant O(1) complexity. The explicit topological coupling constraint, based on predicting contour collisions between adjacent level sets, is developed to further prevent false merging or absorption of neighboring cells, and also reduce fragmentation during level set evolution. The proposed four-color level set algorithm is used to efficiently and accurately segment hundreds of individual epithelial cells within a moving monolayer sheet from time-lapse images of *in vitro* wound healing without any false merging of cells.

1 Introduction

High-throughput content screening using cell image-based assays offers a powerful new tool for understanding the chemical biology of complex cellular processes and offers opportunities for identifying new targets in drug discovery [1]. Image-based live-cell assay experiments need to image and analyze hundreds of thousands of images collected over a short period of time using automated high speed microscopy data acquisition [2]. Since tens of thousands of cells typically need to be screened, highly reliable image analysis algorithms are of critical importance. One fundamental task of automated screening systems is accurate cell segmentation that often precedes other analyses such as cell morphology, tracking and behavior. Since cells undergo complex changes during the cell division process, identifying and segmenting hundreds of closely interacting cells per frame is a challenging task. Cells are also densely clustered due to spatiotemporal sampling

* This work was supported by a U.S National Institute of Health NIBIB award, R33 EB00573.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 101–108, 2006.

and instrumentation cost trade-offs. The accurate, scalable, and computationally efficient segmentation of closely grouped cells, without distinct edges, that approach, touch or overlap each other and whose nuclei become more indistinct at the start of mitosis is the focus of this paper.

Level set based image segmentation techniques [3], in comparison to other algorithms, are well suited to segment a large (unknown) number of deformable but characteristically similar objects (in terms of intensity variation), like cells. Level set methods, adapted to image sequences, are versatile and able to readily adapt to indistinct cell boundaries, appearing or disappearing cells, complex cell shape changes over time, global illumination or focus changes, background motion and non-stationary noise processes.

The two-phase Chan and Vese level set algorithm which segments an image into two sets of possibly disjoint regions, by minimizing a simplified Mumford-Shah functional [4], is ideal for the segmentation of cell image sequences or other sequences in which the intensities or textures of all foreground objects in the images are nearly the same [5]. However, if multiple cells are overlapping, touching or in close proximity to each other, level set methods will tend to merge adjacent contours into a single object (as shown in Fig. 1), which leads to difficulties when tracking individual cells. In order to prevent the merging of cells during tracking, an *implicit* pair-wise region-based coupling constraint was introduced within the level set and parametric snake-based segmentation frameworks by Zhang *et al.,* [6], Dufour *et al.,* [7] and Zimmer and Olivo-Marin [8]. The critical *a priori* assumption that cells do not merge (which is mostly accurate) is used to guide the curve evolution process during segmentation. Without this coupling constraint, cells that are properly segmented in previous frames may get merged with other cells in subsequent frames. The proposed graph vertex coloring (four-color) level set method uses implicit active contours, so we focus our comparison to the coupled geometric level set approach [6].

In [6], the first frame of the image sequence is segmented into $N-$connected components. Each object (cell) is then assigned a unique level set leading to an $N-$level set paradigm. These level sets are subsequently used to track the cells throughout the image sequence. The energy functional, $E_{nls}(\boldsymbol{c_{in}}, c_{out}, \boldsymbol{\Phi})$, used to evolve $N-$ coupled level sets, with $\eta = 0$, is shown below [6]:

$$
\begin{aligned}
E_{nls}(\boldsymbol{c_{in}}, c_{out}, \boldsymbol{\Phi}) &= \mu_{in} \sum_{i=1}^{N} \int_{\boldsymbol{\Omega}} (I - c_{in}^{i})^2 \, H(\phi_i) \, d\boldsymbol{y} \\
&+ \mu_{out} \int_{\boldsymbol{\Omega}} (I - c_{out})^2 \prod_{\substack{i=1 \\ \forall\, i: H(\phi_i) < 0}}^{N} \Big(1 - H(\phi_i)\Big) \, d\boldsymbol{y} + \nu \sum_{i=1}^{N} \int_{\boldsymbol{\Omega}} |\nabla H\big(\phi_i\big)| \, d\boldsymbol{y} \\
&+ \gamma \sum_{i=1}^{N} \sum_{j=i+1}^{N} \int_{\boldsymbol{\Omega}} H(\phi_i) H(\phi_j) \, d\boldsymbol{y} + \eta \Big\{ \sum_{i=1}^{N} \int_{\boldsymbol{\Omega}} \frac{1}{2} (|\nabla \phi_i| - 1)^2 \, d\boldsymbol{y} \Big\} \qquad (1)
\end{aligned}
$$

Here, $\boldsymbol{\Phi} = \{\phi_1, \phi_2, \ldots, \phi_N\}$ represents $N-$level sets associated with N cells in the image; $\boldsymbol{c_{in}}$ represents the average intensities of cells for $\phi_i \geq 0$ while c_{out}

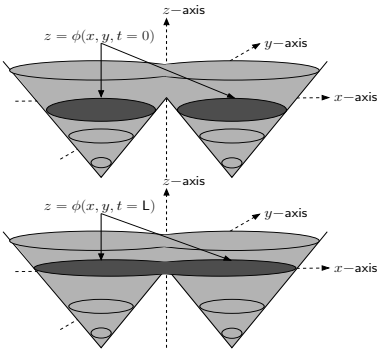


Fig. 1. Neighboring regions get merged with a single level set

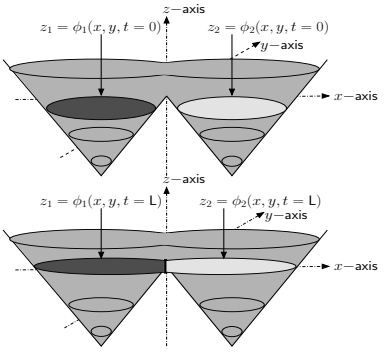


Fig. 2. Topology of neighboring regions preserved with two coupled level sets

is the average intensity of the background[1]. The first and second terms are homogeneity measures of the foregrounds and background of all level sets. The third term controls the lengths of interfaces $\phi_i = 0$, and minimizes the length of all level sets. The fourth term of the functional penalizes pair-wise couplings or overlaps between level sets, while the last term enforces the constraint of $|\nabla\phi_i| = 1$, thus helping us avoid explicit redistancing of level sets during the evolution process [9]. The constants $\mu_{in}, \mu_{out}, \nu, \gamma$, and η are weights associated with each of the terms.

However, the coupled level set algorithm of Zhang *et al.,* is computationally expensive and scales as the square of the number of objects (cells) in the frame. Since each frame may have hundreds to thousands of cells, speed of the algorithm for high-throughput screening studies is extremely important. Trying to evolve thousands of level sets by solving thousands of separate partial differential equations with millions of coupling interaction terms is currently impractical or too expensive even using specialized hardware or parallel processing computing clusters. The number of level sets can be reduced from N to $\lceil \log_2 N \rceil$ using a multiphase Vese and Chan level set segmentation algorithm, as described in [10]. However, there appears to be no mechanism to define spatial coupling constraints between the multiple phases. Each phase segments a set of objects with similar gray-level intensities. So, if two cells touch or overlap each other they will be associated with the same phase, and hence would get merged.

2 Four-Color Level Set Algorithm with Coupling Constraints

In order to improve computational efficiency, we want to take advantage of the fact that objects of interest (i.e., cell nuclei) have similar intensity characteristics. In addition, each object has only a limited number of neighbors compared

[1] The region *exterior to all level sets* indicates the background.

to the total number of objects in the frame. We want to divide the N objects into k *independent* sets (or k-colors) such that objects within each set are not neighbors, and for a given object its clique of neighbors do not share the color of the given object (i.e., are in different independent sets). Hence, in order to prevent cell merges, the first task is to assign cells into a small number of independent sets (i.e., separate colors or level sets). In this scenario:

> *What is the minimum number of "colors", k, that need to be assigned, so that no two neighboring cells have the same "color" ?*

Representing a segmentation as a graph in which cells are vertices, and the adjacency relationships are edges, the problem can be computationally solved using graph-vertex coloring (i.e., graph chromaticity or k-coloring). For general graphs finding the minimum value of k is an NP-complete problem. However, for planar graphs, the famous "Four-Color Theorem" states that at most $k = 4$; any planar graph can be colored with at most four colors such that no two neighboring vertices are assigned the same color [11,12,13]. Although for some graphs $k = 2$, or 3, for most applications including biological cell segmentation typically $k = 4$. Since four level sets are sufficient for 2D cell imaging studies, we set $k = 4$ and do not search for a graph's chromatic number.

Hence, we can now use Eq. 1, with $N = 4$ and make the coupled level set functional *independent* of, N, the number of objects (i.e., cells) in the image. This reduces the number of coupling terms from $O(N^2)$ to $O(1)$ (i.e., a constant number of six coupling terms, per level set, in the energy functional shown in Eq. 1).

Using the realistic *a priori* assumption that all cells in the image have very similar characteristics, enables us to assign a single average intensity c_{in} to all of them (i.e., $\forall\, i,\ c_{in}^{i} = c_{in}$). This is identical to the two-phase Chan and Vese level set algorithm with just one foreground [4], and more efficient than [6] which requires computing N average intensities.

The four Euler-Lagrange evolution equations associated with the minimization of Eq. 1 are as follows ($i = 1, 2, 3, 4$):

$$\frac{\partial \phi_i}{\partial t} = \delta(\phi_i)\Bigg\{ \mu_{in}(I - c_{in}^{i})^2 - \mu_{out}(I - c_{out})^2 \prod_{\substack{j=1 \\ \forall\, j: H(\phi_j)<0, j\neq i}}^{4} \Big(1 - H(\phi_j)\Big) - \nu\, \mathrm{div}\Big(\frac{\nabla \phi_i}{|\nabla \phi_i|}\Big) + \gamma \sum_{j=1; j\neq i}^{4} H(\phi_j) \Bigg\} + \eta \Bigg\{ \Delta\, \phi_i - \mathrm{div}\Big(\frac{\nabla \phi_i}{|\nabla \phi_i|}\Big) \Bigg\} \qquad (2)$$

where, Δ is the Laplacian operator. In our implementation of the energy functional shown in Eq. 1, we use regularized Heaviside and Dirac-delta functions [4].

Computing a single average c_{in} for all objects helps us in randomly associating cells with different level set functions, subject to the four-color criterion being satisfied. Thus, cells associated with a "red" level set in a previous frame, can be associated with a "blue" level set in a subsequent frame. It is necessary to "re-color" the cells at the beginning of the iteration process, as positions of cells may change during the evolution process.

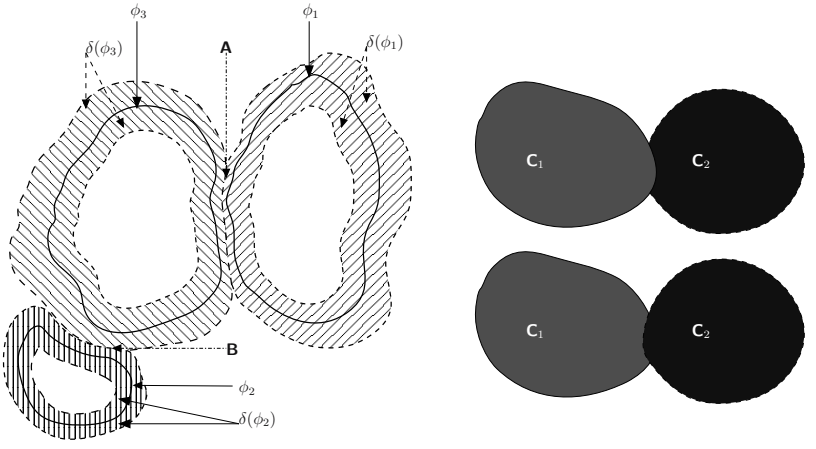


Fig. 3. [**Left**]Narrow-band fronts of cells belonging to three different "colors". The corresponding zero level set curves are given by ϕ_1 ("red"), ϕ_2 ("green"), and ϕ_3 ("blue"), with the narrow bands given by the region where $\delta(\phi_i) > s_{thresh}$. At region **A**, the narrow band of the "red-colored" cell intersects with the narrow band of the "blue-colored" cell. We enforce an explicit rule that the narrow band of a cell, ϕ_i that is currently being updated *cannot overlap* with a different colored cell (i.e., (ϕ_j)). Hence, in region **A**, the narrow-band of the "red-colored" cell (i.e., $\delta(\phi_1) > s_{thresh}$) does not include those areas for which $\phi_3 \geq 0$. A similar rule is applied when updating the "blue-colored" cell, at region **B**. [**Right**] If a "collision" occurs between cells (i.e., $\delta(\phi_i) = \delta(\phi_j), i \neq j$), then depending on which order the level sets are updated, we may get either C_1 protruding into C_2 or vice-versa.

In addition to the energy-based coupling term of Eq. 1 to penalize overlaps between level sets, we also use an explicit topological technique to penalize such overlaps. First, we compute $\delta(\phi_i); i \in [1, 4]$. As we use a narrow-band approach (i.e., $\delta(\phi_i) > s_{thresh}$) to update the level set curves, we check the saliency of $\delta(\phi_i)$ i.e., $\delta(\phi_i) > \delta(\phi_j); j \neq i$ to identify pixels on the front of the current level set that overlaps the narrow-band fronts of other level sets. A pixel on the front of a current level set is updated *only if* its saliency is highest. Topological "collisions" between adjacent cells are detected in this manner and the evolution of the associated level sets near the colliding fronts are stopped (see Fig. 3).

The four-color level set segmentation algorithm is given below:

1. Initialize the segmentation process using level lines [14], and applied to the starting image. Isolate cells that may be touching or very close to each other to produce the segmentation mask (used in step 4).
2. Apply steps 3 to 8 for each frame $f = 1 \ldots N$
3. To speed up convergence project the segmentation mask (i.e., converged level set) from the previous frame as an initial estimate (c.f. [14, 15]).
4. Use the *binary* segmentation mask to extract connected components, their centroids, and associated (planar) adjacency or neighborhood graph such as the Delaunay triangulation (c.f. [16]). Apply a graph-vertex coloring algo-

rithm (c.f. [17]) to partition the cells into four independent sets and produce a *colored* segmentation mask.
5. Associate one level set function with each mask color, and calculate the signed distance functions for each of the four initial zero level sets (i.e., ϕ_i^0 at evolution time $t = 0$).
6. For each frame apply K iterations, and
   (a) Update $\boldsymbol{c_{in}}$ and c_{out};
   (b) Evolve the level set within the narrow band of a cell using Euler-Lagrange equations;
   (c) Enforce an explicit coupling rule that the narrow band of a cell, ϕ_i *cannot overlap* with the level set of any of its neighbors.
7. Generate a binary mask from the four-color segmentation and apply morphological filtering to remove spurious fragments.
8. Apply a spatially-adaptive level line-based coarse segmentation to the background (i.e., complement of the dilated colored segmentation mask), in order to detect objects not present in the previous frame (i.e., new objects entering the current frame).

3 Results and Analysis

The proposed algorithm was tested on a wound healing image sequence consisting 136 frames, with dimensions of 300×300 pixels (40μm×40μm). The sequence used in our simulations was obtained using a monolayer of cultured porcine epithelial cells, as described by Salaycik *et al.,* in [18]. Images were sampled uniformly over a 9:00:48 hour period and acquired using phase contrast microscopy with a 10× objective lens, at a resolution of approximately 0.13μm per pixel. We provide two sets of results for our four-color level set algorithm with $\mu_{in} = 1, \mu_{out} = 1, \nu = 0.004, \gamma = 1.0, \eta = 0.1$, while the number of iterations K is set to 10 for each frame. A qualitative comparison in Fig. 4 shows the benefit of an explicit topological coupling constraint on two representative frames (113 and 124) where cell merges are correctly prevented. A quantitative comparison in Table 1 shows the benefit of using the (energy-based and explicit) coupling constraints in preventing 46 manually verified cell merge events. Splits, merges, appearances (App.) and disappearances (Disapp.) in Table 1 were confirmed manually, since complete ground truth for such a complex sequence is difficult and time consuming to construct. Appearances and disappearances indicate cells that are not associated with cells in the previous or next frame, respectively (i.e., cells entering or leaving the frame, or cell apoptosis). Splits and merges are cells that can be associated with multiple parents or children in the trajectory, respectively (i.e., cell mitosis or cell clumping). However, segmentation artifacts or tracking mis-associations can produce any of these events. Cell splits, merges, appearances or disappearances can be due to biological events, or segmentation and tracking errors [5]. A split-merge-split cycle which is indicative of fragmentation during segmentation leads to a high score for split and merge events. The results shown in Table 1, confirm the advantage of using coupling in our four-color level set algorithm.

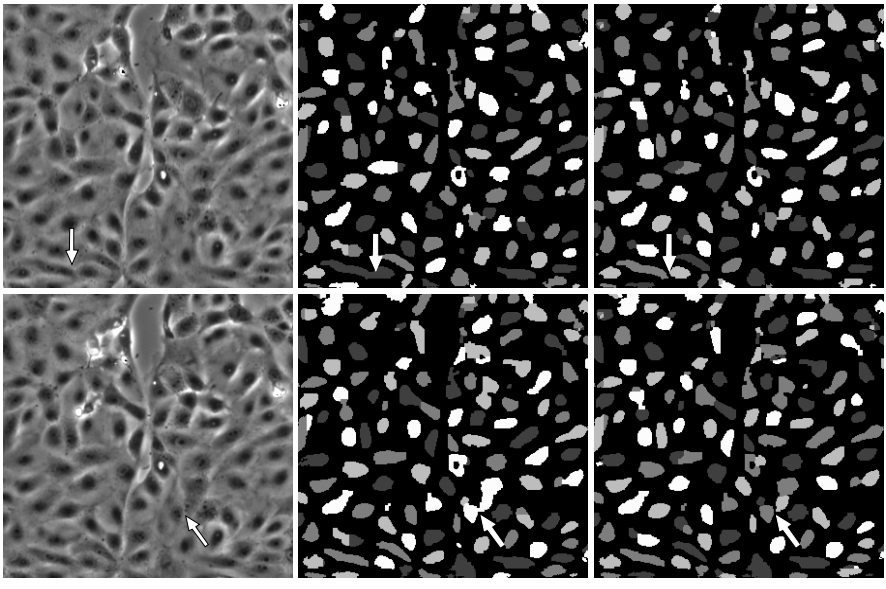


Fig. 4. The left column indicates frame numbers 113 and 124 from the wound healing image sequence. The central column depicts segmentation results when our four level-set functional, without any explicit control on the evolution of level set functions. The right column indicates segmentation results when using our four level-set algorithm, with explicit coupling. The "color" masks are different in each frame due to a "re-coloring" process applied at the beginning of the iteration process. Images have been scaled for display purposes. The arrows in the central and right columns show where cell merge events occur.

Table 1. Tracking results, when using the four-color level set algorithm, with and without coupling. With coupling indicates both energy-based ($\gamma = 1.0$) and explicit coupling constraints. Without coupling indicates ($\gamma = 0$) as well as no explicit coupling.

	Objects	**Splits**	**Merges**	**App.**	**Disapp.**
`Uncoupled`	13986	73	46	24	35
`Coupled`	14570	34	0	21	30

4 Conclusions

A novel four-color level set algorithm for segmenting N cells (objects) based on graph-vertex coloring was presented, using the "active contour without edges" level set coupled energy functional, combined with a new explicit topological object to object coupling constraint. The four color level set algorithm dramatically reduces the computational cost of incorporating coupling constraints to prevent cells (objects) from merging, from $O(N)$ level sets and $O(N^2)$ coupling constraints to $O(1)$ level sets and $O(1)$ coupling constraints for N objects. The reduction in the number of level sets (N to 4) and energy-based coupling constraints in the

Euler-Lagrange equations ($N^2 - N$ to 12), with 12 explicit topological checks per pixel makes the proposed algorithm highly scalable, and computationally resource efficient for segmenting a large number of complex shaped objects.

References

1. M.M.Morelock, *et al.,* : Statistics of assay validation in high throughput cell imaging of nuclear factor κB nuclear translocation. *ASSAY Drug Dev. Tech.* **3** (2005) 483–499
2. B.Neumann, *et al.,* : High-throughput RNAi screening by time-lapse imaging of live human cells. *Nature Methods* **3** (2006) 385–390
3. J.A.Sethian: *Level Set Methods and Fast Marching Methods*. Cambridge University Press, New York, USA (1999) ISBN 0-521-64557-3.
4. T.Chan, L.Vese: Active contours without edges. *IEEE Trans. Image Process.* **10** (2001) 266–277
5. F.Bunyak, K.Palaniappan, S.K.Nath, T.I.Baskin, G.Dong: Quantitive cell motility for *in vitro* wound healing using level set-based active contour tracking. In: *Proc. 3rd IEEE Int. Symp. Biomed. Imaging (ISBI)*, Arlington, VA (2006) 1040–1043
6. B.Zhang, C.Zimmer, J.-C.Olivo-Marin: Tracking fluorescent cells with coupled geometric active contours. In: *Proc. 2nd IEEE Int. Symp. Biomed. Imaging (ISBI)*, Arlington, VA (2004) 476–479
7. A.Dufour, V.Shinin, S.Tajbakhsh, N.Guillén-Aghion, J.-C.Olivo-Marin, C.Zimmer: Segmenting and tracking fluorescent cells in dynamic 3-D microscopy with coupled active surfaces. *IEEE Trans. Image Process.* **14** (2005) 1396–1410
8. C.Zimmer, J.-C.Olivo-Marin: Coupled parametric active contours. *IEEE Trans. Pattern Anal. Machine Intell.* **27** (2005) 1838–1842
9. C.Li, C.Xu, C.Gui, D.Fox: Level set evolution without re-initialization: A new variational formulation. In: *Proc. IEEE Conf. Computer Vision Pattern Recognition.* Volume 1. (2005) 430–436
10. L.Vese, T.Chan: A multiphase level set framework for image segmentation using the Mumford and Shah model. *Intern. J. Comput. Vis.* **50** (2002) 271–293
11. K.Appel, W.Haken: Every planar map is four colorable. Part I. discharging. *Illinois. J. Math.* **21** (1977) 429–490
12. K.Appel, W.Haken, J.Koch: Every planar map is four colorable. Part II. reducibility. *Illinois. J. Math.* **21** (1977) 491–567
13. N.Robertson, D.P.Sanders, P.D.Seymour, R.Thomas: The four color theorem. *J. Combin. Theory, Ser. B* **70** (1997) 2–44
14. D.P.Mukherjee, N.Ray, S.T.Acton: Level set analysis for leukocyte detection and tracking. *IEEE Trans. Image Process.* **13** (2001) 562–672
15. N.Paragios, R.Deriche: Geodesic active contours and level sets for the detection and tracking of moving objects. *IEEE Trans. Pattern Anal. Machine Intell.* **22** (2000) 266–280
16. G.Seetharaman, G.Gasperas, K.Palaniappan: A piecewise affine model for image registration in non-rigid motion analysis. In: *Proc. IEEE Int. Conf. Image Processing.* Volume 1., Vancouver, Canada (2000) 561–564
17. D.Brèlaz: New methods to color the vertices of a graph. *Comm. ACM* **22** (1979) 251–256
18. K.J.Salaycik, C.J.Fagerstrom, K.Murthy, U.S.Tulu, P.Wadsworth: Quantification of microtubule nucleation growth and dynamics in wound-edge cells. *J. Cell Sci.* **118** (2005) 4113–4122

3D Histological Reconstruction of Fiber Tracts and Direct Comparison with Diffusion Tensor MRI Tractography

Julien Dauguet[1], Sharon Peled[2], Vladimir Berezovskii[3], Thierry Delzescaux[4], Simon K. Warfield[1], Richard Born[3], and Carl-Fredrik Westin[5]

[1] Computational Radiology Laboratory, Children's Hospital, Brigham and Women's Hospital, Harvard Medical School, Boston, USA
[2] Harvard Center for Neurodegeneration and Repair, Boston, USA
[3] Department of Neurobiology, Harvard Medical School, Boston, USA
[4] Service Hospitalier Frédéric Joliot, CEA, Orsay, France
[5] Laboratory of Mathematics in Imaging, Brigham and Women's Hospital, Harvard Medical School, Boston, USA

Abstract. A classical neural tract tracer, WGA-HRP, was injected at multiple sites within the brain of a macaque monkey. Histological sections of the labeled fiber tracts were reconstructed in 3D, and the fibers were segmented and registered with the anatomical post-mortem MRI from the same animal. Fiber tracing along the same pathways was performed on the DTI data using a classical diffusion tracing technique. The fibers derived from the DTI were compared with those segmented from the histology in order to evaluate the performance of DTI fiber tracing. While there was generally good agreement between the two methods, our results reveal certain limitations of DTI tractography, particularly at regions of fiber tract crossing or bifurcation.

1 Introduction

Tracing neural connections lies at the heart of neuroanatomy, and has profound implications for the study of neural function and for the study of developmental and adult plasticity of the nervous system. It is also important for the study of neurodegenerative diseases, and for the planning of neurosurgical procedures, such as tumor ablation. Methods for tracing connections in the brain have a long history, beginning with those based on lesions and the resulting retrograde or anterograde degeneration [Nissl 1894; Marchi 1886; Nauta and Gygax 1954]. Subsequent methods exploited the axonal transport of specific molecules, beginning with horseradish peroxidase (HRP) [1], which opened up a new era in neuroanatomy. This was followed by the discovery of a host of other tracers including small fluorescent molecules [2,3], lectins [4], neurotrophins [5], neurotoxins [6], dextrans [7], and, more recently, carbocyanine dyes [8], latex microspheres [9] and viruses [3]. While constituting a powerful armamentarium with which to study connections in the brain, all of these methods possess the obvious

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 109–116, 2006.

drawback that they are highly invasive and require sacrifice of the experimental animal for histological visualization of the transported substance.

In the 1970's, water diffusion magnetic resonance imaging was introduced and later used for medical applications [10]. With the introduction of diffusion tensor imaging (DTI) [11], white matter tractography based on DTI became the first non-invasive, in vivo technique for gross fiber tract tracing. Since then, many advanced methods and models have been proposed to improve the quality of the tractography [12,13,14]. It is important to stress the fact that the explicit quantity measured is water diffusion and not fibers. Moreover, since MRI methods in general always obtain a macroscopic measure of a microscopic quantity, which necessarily entails intravoxel averaging, the voxel dimensions influence the measured diffusion tensor at any particular location in the brain.

As the sophistication of DTI tractography algorithms increases, it becomes essential to validate the results. These techniques are already used in many biomedical studies. It is clear that DTI reveals gross connectivity as described in atlases based on histology (e.g. [15]), but, as of yet, no rigorous validation of these techniques has been done. In particular, the performance of DTI tractography is unknown in the case of fiber decussations and the transition of fibers through a relay. Moreover, the ability of the methods to give reliable information on the density of fibers remains an open issue. Here we compare tracts revealed with WGA-HRP on histology to the corresponding tracts calculated from DTI data in a macaque brain. Our group presented preliminary results on this data in [16].

2 Material and Method

Using procedures approved by the Harvard Medical Area Standing Committee on Animals, craniotomies were performed on a Macaque monkey under general anesthesia, and 4% Wheat Germ Agglutinin (WGA) conjugated to HRP (WGA-HRP) in water solution was pressure-injected under microscopic stereotaxic guidance. WGA is a plant lectin that transported in both the retrograde and anterograde directions, and that binds to specific cell-surface glycans, in particular, N-acetylglucosamine and sialic acid, which are then internalized via receptor-mediated endocytosis. Though we use the term in the discussion below, WGA is not a "stain" per se; it is rendered visible by virtue of a reaction with the peroxidase. Four 0.25 microliter injections were made in primary visual cortex (V1), five 0.2 microliter injections each along the pre- and post-central gyri and one 5 microliter injection into the left eye. Three days later the monkey was sacrificed, perfused with aldehyde fixative and the brain removed from the skull.

2.1 MR Image Acquisition

Post-mortem MR imaging of the fixed brain was performed on a 4.7T Bruker Biospec using a 3D gradient echo proton density sequence (TR 100, flip angle 10) for anatomy. Diffusion weighted imaging (DWI) was performed using a spin-echo sequence (30 directions, bvalues of 1000 s/mm^2), with voxel dimensions

$0.5 \times 0.5 \times 1\text{mm}^3$. *In vivo* DWI data was also acquired prior to sacrifice but is not described here.

2.2 Brain Sectioning and Histo-chemistry

After MR imaging, the brain was separated into two blocks to facilitate sectioning. The whole brain was serially sectioned in the coronal plane ($80\mu m$ thickness) using a freezing, sliding microtome. Prior to cutting each section, the block face was photographed using a rigidly mounted Hasselblad camera with a 25 Mpixel Phase 1 digital back. This greatly facilitated subsequent 3D reconstruction.

Every eighth section was reacted for WGA-HRP using tetramethyl benzidine [17], producing a series of 90 sections spanning the entire brain. These sections were then mounted on glass slides and scanned using both a GE Healthcare Amersham Typhoon 9410 in fluorescent transparency mode at 100μm resolution, and a regular flat-bed scanner in reflection optical mode at 400dpi resolution. The fluorescent scanning optimizes the grey/white matter contrast for registration purposes whereas the optical scanning allows differentiation of fiber staining for segmenting the bundles. A linear combination of these two scans produced a hybrid image with both good contrast and clear fiber staining. In this linear combination the optical scanning images were inverted producing white areas of HRP reacted tissue.

2.3 Photographic Volume and Spatial Alignment

The block face photographs corresponding to the subset of stained and mounted histological slices were selected and stacked. Within each of the anterior and posterior blocks, the stacked photographs were aligned slice by slice with the flat-bed scanned histology sections. The anterior and posterior blocks were then assembled together using an affine registration. The resulting image data set will be referred to as the "photographic volume."

Each histological slice was then co-aligned on the corresponding photograph using per-hemisphere affine registration, consist of automatically segmenting the hemispheres and estimating an individual 2D affine transformation for each one [18]. This volume of histological slices will be referred to as the histological volume. A 3D affine tranformation followed by an elastic Free Form Deformation (FFD) [19] were estimated between the photographic volume and the anatomical MRI data to compensate the mechanical deformations due to the handling of the brain during the two scans [20]. These transformations were applied to the histological volume to get both MRI data and histological volume in the same geometry.

2.4 Tracing DTI and Histological Volume

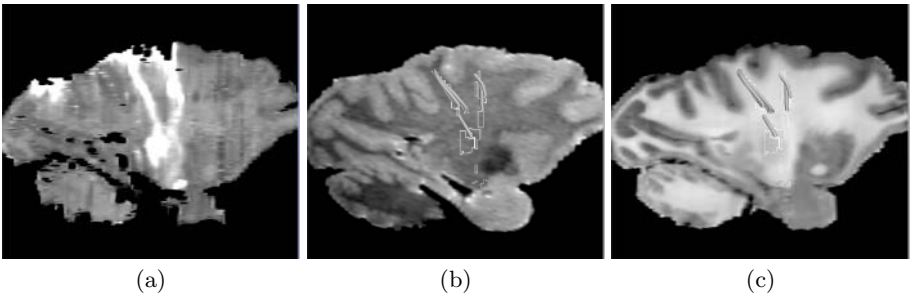
The HRP reacted areas in the histological volume were manually segmented. This was performed by tracing them through the histological slices, and refined by inspecting reformatted orthogonal views.

The fiber segmentation from the histological volume then was used to as a guide to define seed points for the tractography performed in the DTI volume. A standard hyper-streamline based method for tractography was used [14].

3 Results

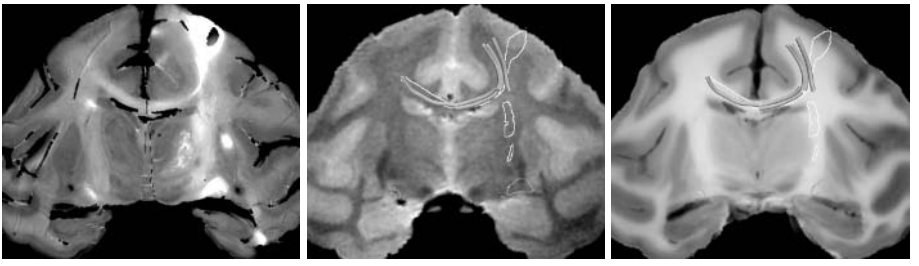
The three different data sets (the DTI data volume, the photographic volume, and the HRP stained histological volume) were spatially aligned to a common coordinate system to allow for direct comparison between results from DTI tractography and the HRP staining. The histological and photographic volumes were registered to the MRI geometry. Segmentations of fibers according to the staining on the histological slices and the tracts derived from the DTI are presented for comparison for all the injection sites.

In the rest of this section we present five figures, the first four of which show cross-sections of the HRP reacted volume and the segmented fibers overlayed in the structural MRI and photographical volumes. Figure 1 shows sagittal views of the segmented fibers connecting the primary motor cortex (precentral gyrus) and the pons via the internal capsule, along with the fibers connecting the somatosensory cortex (postcentral gyrus) and the ventral posterio-lateral nucleus (VPL) of the thalamus. Figure 2 shows fibers in the primary motor cortex (same injection as shown in Figure 1), connecting the left and right hemispheres via the corpus callosum. Figure 3 shows fibers connecting the primary visual cortex (V1) and the lateral geniculate nucleus (LGN) in the thalamus. Figure 4 shows fibers connecting the left eye to the lateral geniculate nucleus (LGN) in the thalamus via the optic chiasm (the optic nerves can not be seen because they are outside of the brain). Figure 5 shows the combined result of the histological tract segmentation and the DTI tractography from Figures 1-4.



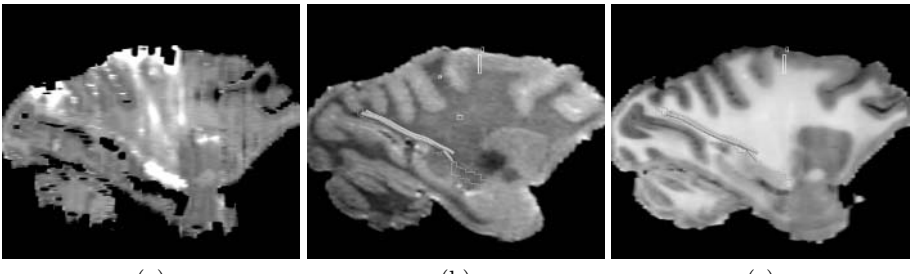
(a) (b) (c)

Fig. 1. Fibers connecting the primary motor cortex (precentral gyrus) to the pons via the internal capsule and fibers connecting the somatosensory cortex (postcentral gyrus) from the ventral posterio-lateral nucleus (VPL) of the thalamus. The figure shows (a) sagittally reformatted view of the HRP reacted histological volume (bright in this picture), (b) fiber pathways derived from histology DTI tractography superimposed on the corresponding structural MRI data and (c) on the photographic volume.



(a) (b) (c)

Fig. 2. Fibers in the primary motor cortex (precentral gyrus) connecting the hemispheres via the corpus callosum. The figure shows (a) coronal view of the HRP reacted histological volume (bright in this picture), (b) fiber pathways derived from histology DTI tractography superimposed on the corresponding structural MRI data and (c) on the photographic volume.



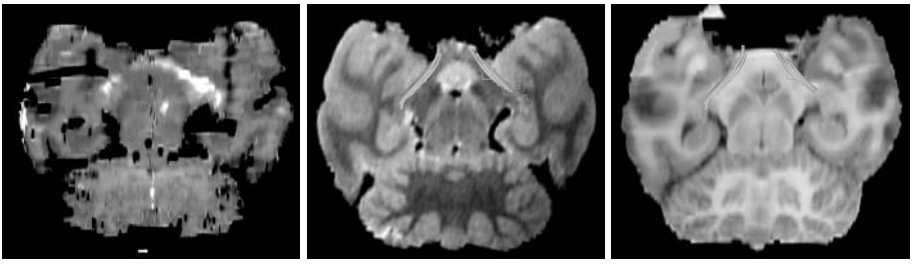
(a) (b) (c)

Fig. 3. Fibers in the primary visual cortex (V1) connecting to the lateral geniculate nucleus (LGN) in the thalamus. The figure shows (a) sagittal view of the HRP reacted histological volume (bright in this picture), (b) fiber pathways derived from histology DTI tractography superimposed on the corresponding structural MRI data and (c) on the photographic volume.

4 Discussion

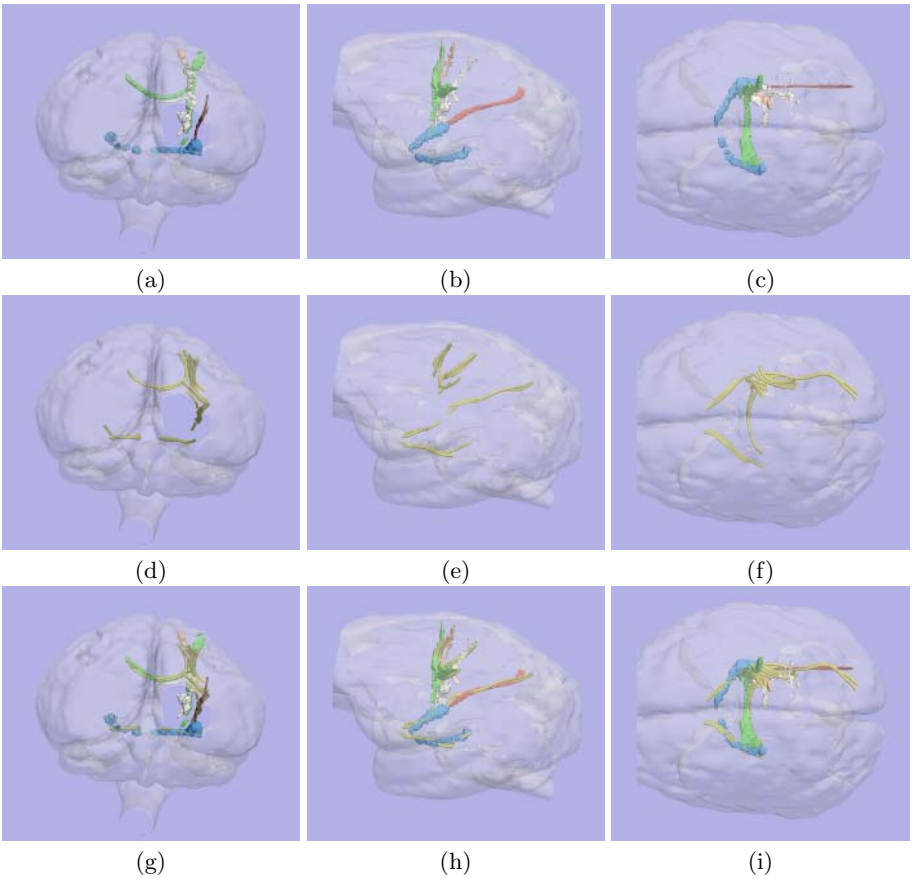
In this paper we have presented results visually comparing tracts derived from DTI tractography and tracts derived from histology sections of a macaque brain. To enable direct comparisons, the data sets were spatially aligned to a common coordinate system using non-linear registration methods. To the best of our knowledge this study is the first that compares long range connectivity derived from histology in 3D and results from DTI tractography.

The results presented show that the tracts derived from DTI correlate well visually with the corresponding tracts derived from the HRP reacted histology volume, as shown in the chiasma (Figure 4) and the connections to the visual cortex (Figure 3). However, some discrepancies can be noticed. For the tract seed points in isotropic areas such as in the cortex, the DTI tractography fails



(a) (b) (c)

Fig. 4. Fibers connecting the left eye to the lateral geniculate nucleus (LGN) in the thalamus via the chiasma. The figure shows (a) axial view of the HRP reacted histological volume (bright in this picture), (b) fiber pathways derived from histology DTI tractography superimposed on the corresponding structural MRI data and (c) on the photographic volume.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Fig. 5. Three different views of the paths derived from histological slices (a, b, c), the corresponding fiber tracts derived from DTI (d, e, f), and both results displayed together (g, h, i)

to trace the main bundles. Figure 1) shows that tracts derived from DTI did not continue into the cortex injection sites.

In general, DTI tractography tends to have difficulties when tracts cross or divide. When reaching such areas, the tracts often stop due to low anisotropy of the diffusion tensors or fail to follow the correct branch when tracts divide. This was noticed for the pathways from the primary motor cortex injection going both to the corpus callosum and to the internal capsule. The DTI tractography from the motor cortex injection, shown in Figure 1 and in Figure 2, could not be used to trace both pathways, and additional seed points along the segmented tract were needed. The problem of terminated tracts was noticed when fibers, which were seeded from the optic tract, stopped in the lateral geniculate nucleus (Figures 3 and 4). The resolution of crossing fibers is a known problem in tractography - modeling of two or more tensors in each voxel is an emerging research area that may help in solving this problem [21,22,23,24]. Future work will include studies on how to optimize parameter settings for tractography methods, and how to incorporate more complex diffusion models than the single tensor model to better handle complex fiber regions in DTI tractography.

Acknowledgements. The authors gratefully acknowledge Dr. Phillip Hendrickson for his help with the injection and the imaging, and the support of NIH grants P41-RR13218, R01-MH50747, and P30-EY12196. This investigation was also supported in part by NSF ITR 0426558, grant RG 3478A2/2 from the NMSS, and by NIH grants R21 MH67054, P41 RR013218 and U41 RR019703. This work is part of the National Alliance for Medical Image Computing (NAMIC), funded through the NIH Roadmap for Medical Research, Grant U54 EB005149. Information on the National Centers for Biomedical Computing can be obtained from http://nihroadmap.nih.gov/bioinformatics.

References

1. La Vail, J., La Vail, M.: Retrograde axonal transport in the central nervous system. Science **176** (1972) 1416–1417
2. Bentivoglio, M., Kuypers, H., Catsman-Berrevoets, C., Loewe, H., Dann, O.: Two new fluorescent retrograde neuronal tracers which are transported over long distances. Neurosci Lett. **18**(1) (1980) 25–30
3. Kuypers, H., Ugolini, G.: Viruses as transneuronal tracers. Trends Neurosci. **13**(2) (1990) 71–75
4. Schwab, M., Javoy-Agid, F., Agid, Y.: Labeled wheat germ agglutinin (wga) as a new, highly sensitive retrograde tracer in the rat brain hippocampal system. Brain Res. **152** (1978) 145–150
5. Hendry, I., Stockel, K., Thoenen, H., Iversen, L.: The retrograde axonal transport of nerve growth factor. Brain Res. **68** (1974) 103–121
6. Stoeckel, K., Schwab, M., Thoenen, H.: Role of gangliosides in the uptake and retrograde axonal transport of cholera and tetanus toxin as compared to nerve growth factor and wheat germ agglutinin. Brain Res. (1977)

7. Glover, J., Petursdottir, G., Jansen, J.: Fluorescent dextran-amines used as axonal tracers in the nervous system of the chicken embryo. J Neurosci Methods. **18** (1986) 243–254
8. Honig, M., Hume, R.: Fluorescent carbocyanine dyes allow living neurons of identified origin to be studied in long-term cultures. J Cell Biol. **103** (1986) 171–187
9. Katz, L., Burkhalter, A., Dreyer, W.: Fluorescent latex microspheres as a retrograde neuronal marker for in vivo and in vitro studies of visual cortex. Nature **310** (1984) 498–500
10. LeBihan, D., Breton, E., Lallemand, D.: MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. Radiology **161** (1986) 401–407
11. Basser, P., Mattiello, J., LeBihan, D.: MR diffusion tensor spectroscopy and imaging. Biophys. J. **66**(1) (1994) 259–267
12. Mori, S., Crain, B., Chacko, V., van Zijl, P.: Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. Ann Neurol. **45** (1999)
13. Basser, P.J., Pajevic, P.J., Pierpaoli, C., Duda, J., Aldroubi, A.: In vivo fiber tractography using dt-mri data. Magn. Reson. Med. (44) (2000) 625–632
14. Westin, C.F., Maier, S., Mamata, H., Nabavi, A., Jolesz, F., Kikinis, R.: Processing and visualization of diffusion tensor MRI. Medical Image Analysis **6** (2002) 93–108
15. Nolte, J.: The Human Brain. An Introduction to Its Functional Anatomy. C.V. Mosby; 5th edition (2002)
16. Peled, S., Berezovskii, V., Hendrickson, P., Born, R., Westin, C.F.: Histological validation of DTI using WGA-HRP in a macaque. In: Proc. ISMRM, Miami (2005)
17. Mesulam, M.: Tetramethyl benzidine for horseradish peroxidase neurohistochemistry: a non-carcinogenic blue reaction product with superior sensitivity for visualizing neural afferents and efferents. J. Histochem Cytochem. **26** (1978)
18. Dauguet, J.: L'imagerie post mortem tridimensionnelle cérébrale. PhD thesis, École Centrale de Paris (2005)
19. 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 Transactions on Medical Imaging **18**(8) (1999) 712–721
20. Delzescaux, T., Dauguet, J., Condé, F., Maroy, R., Frouin, V.: Using 3D non rigid FFD-based method to register post mortem 3D histological data and in vivo MRI of a baboon brain. In: MICCAI 2003. Volume 2879 of LNCS. (2003) 965–966
21. Tuch, D.S.: Diffusion MRI of Complex Tissue Structure. PhD thesis, MIT (2002)
22. Perrin, M., Poupon, C., Rieul, B., Leroux, P., Constantinesco, A., Mangin, J.F., Bihan, D.L.: Validation of q-ball imaging with a diffusion fibre-crossing phantom on a clinical scanner. Phil. Trans. R. Soc. B **360** (2005) 881–891
23. Peled, S., Westin, C.F.: Geometric extraction of two crossing tracts in DWI. In: Proc. ISMRM, Miami (2005)
24. Bergmann, Ø., Kindlmann, G., Lundervold, A., Westin, C.F.: Diffusion k-tensor estimation from q-ball imaging using discretized principal axes. In: MICCAI'06. Lecture Notes in Computer Science, Copenhagen, Denmark (2006)

Rician Noise Removal in Diffusion Tensor MRI

Saurav Basu, Thomas Fletcher, and Ross Whitaker

University of Utah,
School of Computing,
Salt Lake City, UT 84112, USA

Abstract. Rician noise introduces a bias into MRI measurements that can have a significant impact on the shapes and orientations of tensors in diffusion tensor magnetic resonance images. This is less of a problem in structural MRI, because this bias is signal dependent and it does not seriously impair tissue identification or clinical diagnoses. However, diffusion imaging is used extensively for quantitative evaluations, and the tensors used in those evaluations are biased in ways that depend on orientation and signal levels. This paper presents a strategy for filtering diffusion tensor magnetic resonance images that addresses these issues. The method is a maximum *a posteriori* estimation technique that operates directly on the diffusion weighted images and accounts for the biases introduced by Rician noise. We account for Rician noise through a data likelihood term that is combined with a spatial smoothing prior. The method compares favorably with several other approaches from the literature, including methods that filter diffusion weighted imagery and those that operate directly on the diffusion tensors.

1 Introduction

The quality of DT-MRI images is limited by the relatively long acquisition times necessary to obtain data at high spatial resolutions. Because acquisition time is restricted by issues of patient comfort and system demand, the signal-to-noise ratio(SNR) in DT-MRI is often low. Thus, post processing techniques to remove noise in the acquired data are important. The diffusion weighted images (DWIs), from which the tensors are derived, are corrupted by Rician noise, which introduces a positive bias in those measurements. These signal-dependent biases are not so detrimental to structural imaging, because they typically do not interfere with diagnostic decisions or tissue classification. However, DT-MRI measurements are being used extensively for quantitative comparisons, and several studies [1,2,3] have shown that bias can affect tensor properties such as trace and fractional anisotropy (FA).

Previous DT-MRI filtering methods fall into two categories: filters that operate on the DWIs and those that operate directly on the tensors. For instance Parker *et al.* [4] use Perona & Malik (P&M) [5] anisotropic diffusion to filter DWIs, and show that it produces less distortion in FA than filtering images of FA. Wang *et al.* [6] formulate a variational approach to regularize DWIs while

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 117–125, 2006.

constraining the estimated tensors to remain positive definite. Martin *et al.* [7] develop a Gaussian Markov Random Field model to regularize the estimated diffusion tensor images. Pennec *et al.* [8] describe a framework for performing anisotropic diffusion on tensors which preserves the property of tensors being symmetric and positive definite. Their filter is based on the idea that the space of all positive definite symmetric matrices forms a Riemannian manifold with each point representing a diffusion tensor. None of these techniques explicitly account for the effects of bias in the original DWI measurements. After submission of this work, we became aware of the work by Fillard et al. [9] that adds a Rician noise model to smoothing of tensor images in a Log-Euclidean framework.

In this paper we show Monte Carlo simulations that add new insights into the effects of Rician bias on tensor measurements. These results demonstrate the need for realistic noise models in DT-MRI filtering. We describe a filtering strategy that explicitly models the Rician noise as a data likelihood term in a maximum *a posteriori* framework. To assess the performance of our technique, we propose a new method for producing low noise DWIs using a maximum likelihood estimate (MLE) from repeated scans of a healthy volunteer. We present a comparison of filtering performance for tensor based methods and methods that smooth the DWIs. Our results show that filtering on the original DW images gives better results than filtering on tensor images, and that our method using an explicit model of Rician noise gives the best overall results.

2 Rician Noise and Its Effects on Diffusion Tensors

It is well known that MR magnitude images are corrupted by Rician noise, which arises from complex Gaussian noise in the original frequency domain (k-space) measurements. The Rician probability density function for the corrupted image intensity x is given by

$$p(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2 + A^2}{2\sigma^2}\right) I_0\left(\frac{xA}{\sigma^2}\right), \tag{1}$$

where A is the underlying true intensity, σ is the standard deviation of the noise, and I_0 is the modified zeroth-order Bessel function of the first kind.

Previous studies on the effect of noise on diffusion tensor measurements have shown that as noise increases, the tensor trace decreases [1] and FA increases [1,2,3]. Here we show that these effects can actually be quite different depending on the orientation of the diffusion tensor with respect to the measurement gradients. Using power series analysis, Anderson [2] shows that the major eigenvalue increases with higher noise, causing FA to increase. This analysis assumes the major eigenvalue is a combination of several diffusion weighted measurements, which happens when the major eigenvector lies in between several gradient directions. However, consider the special case of six gradient directions where the

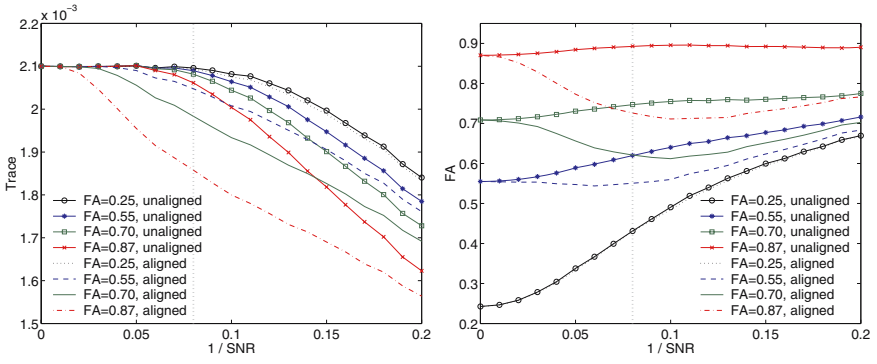


Fig. 1. Results of Monte Carlo simulations for trace (left) and FA (right). As an example, the SNR from the real data in section 4.2 is plotted as a vertical line.

major eigenvector is aligned with one gradient direction. Here the major eigenvalue will be completely determined from the diffusion weighted measurement in that direction. In this case, the Stejskal-Tanner equation for the ith diffusion weighted value reduces to $A_i = A_0 \exp(-b\lambda_1 \|\mathbf{g}_i\|)$, where λ_1 is the major eigenvalue, A_0 is the baseline T2 value, and $\mathbf{g}_i$ is the gradient direction with which the major axis is aligned. Since λ_1 is large, the value A_i will be much lower than the baseline A_0, and will thus be more susceptible to Rician bias. This positive bias will tend to underestimate the diffusion in the major eigendirection. We can thus expect two things: the underestimation effect of the tensor trace will be greater and FA will actually go down.

To test the hypothesis that tensor orientation changes the effects of noise on diffusion tensors, we performed Monte Carlo simulations of diffusion tensors characteristic of those found in brain white matter. These tensors had constant trace of 2.1×10^{-3} mm^2/s and four levels of fractional anisotropy, with eigenvalues $\lambda_1 > \lambda_2 \approx \lambda_3$. We used a b-value of 2000 s mm^{-2} and a commonly used sampling scheme of six gradient directions [10]. We repeated the tests for two tensor orientations: one with major axis $(1, 0, 0)$, equally splitting the surrounding gradient directions, and one with the major axis aligned with a gradient direction. We varied the 1/SNR from 0 to 0.2 and used 10,000 tensors per trial.

Figure 1 shows the average trace and FA of the simulated tensors as a function of the noise level. Beyond what has previously been reported, our results show that the amount of bias in the trace increases when the tensor has higher anisotropy or when the tensor is aligned with a gradient direction. Our results show that tensors that are not aligned with gradient directions tend to have increased FA, while those aligned with a gradient direction tend to have decreased FA for moderate levels of noise. However, for lower b-values (e.g., $b = 1000$) these effects are less pronounced. The fact that FA can be overestimated or underestimated depending on the orientation of the fiber tract in the scanner has serious consequences for clinical studies using DT-MRI.

3 Rician Bias Correction Filter

Our Rician bias correction filter is based on a maximum *a posteriori* (MAP) approach to the image reconstruction problem. Given an initial noisy image u_0, we construct the filtered image u that maximizes the log-posterior probability

$$\log p(u|u_0) \propto \log p(u_0|u) + \log p(u), \tag{2}$$

where $p(u_0|u)$ is the likelihood term, or noise model, and $p(u)$ is the prior. For DWIs we consider u to be a vector-valued image, each gradient direction (including $b = 0$) representing a vector component. The formulation in this section would also be valid for structural MRI.

3.1 The Rician Likelihood Term

The formulation of the filtering problem as maximization of a posterior $p(u|u_0)$ is useful as it allows us to incorporate the Rician bias correction as a data attachment term which can be added to the prior model. Using the Rician distribution (1) as the likelihood term and assuming independent noise, the pointwise log-likelihood becomes

$$\log p(u_0|u) = \log \frac{u_0}{\sigma^2} - \frac{u_0^2 + u^2}{2\sigma^2} + \log I_0 \left(\frac{u_0 u}{\sigma^2} \right). \tag{3}$$

The derivative of (3) with respect to u, gives Rician data attachment term

$$\mathcal{B} = -\frac{u}{\sigma^2} + \left[I_1 \left(\frac{u_0 u}{\sigma^2} \right) / I_0 \left(\frac{u_0 u}{\sigma^2} \right) \right] \frac{u_0}{\sigma^2}. \tag{4}$$

3.2 Combining the Rician Model with a Prior

The data likelihood term can be combined with any image prior model. In this paper we use a Gibb's prior model based on a P&M energy functional, given by

$$p(u) = \frac{1}{z} \exp(-E(u)), \qquad E(u) = \lambda \int_U c \left(\|\nabla u\|^2 \right) \, dx \, dy, \tag{5}$$

where z is a suitable normalization, U is the image domain and c is the conductance given by $c(\|\nabla u\|^2) = \exp(-\|\nabla u\|^2 / 2k^2)$, k is the conductance parameter and λ is a constant weighting term.

By adding the Rician likelihood term (4) with the variational of the P&M energy functional we form the update equation for the filtered image,

$$\frac{\partial u}{\partial t} = \mathcal{B} + \lambda \operatorname{div} \left(c \left(\|\nabla u\|^2 \right) \nabla u \right). \tag{6}$$

4 Results

The performance comparisons were made on four different filtering methods: Euclidean tensor filtering, Riemannian tensor filtering [8] and vector anisotropic diffusion on DWIs with and without the Rician likelihood term as described in section 3.1. Here Euclidean filtering refers to vector anisotropic diffusion on the tensor components. To compare these methods, we used both synthetic and real datasets. We used three different error metrics - root mean squared (RMS) error in the tensor, trace and fractional anisotropy. The error between tensors is computed using the Frobenius norm. The parameters for each method were optimized for the RMS error on the tensor components. Both synthetic and real datasets use seven images for each slice, one without diffusion gradient (b=0) and the remaining six with b=1000s/mm^2 and diffusion gradients along the standard six orientations [10].

4.1 Synthetic Data

We used a 10×10×4 volume of tensors oriented in two directions so as to have a tensor discontinuity. One group of tensors have major axes that split the gradient directions, while the others are aligned with a gradient. Figure 2 shows the clean and noisy synthetic tensors with SNR=15.

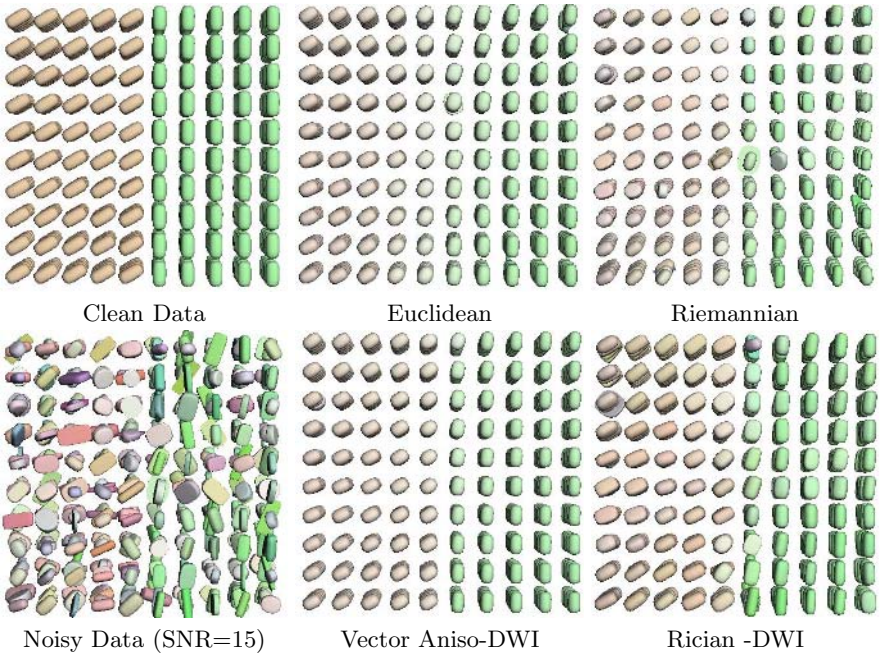


Fig. 2. Synthetic Data Filter Results

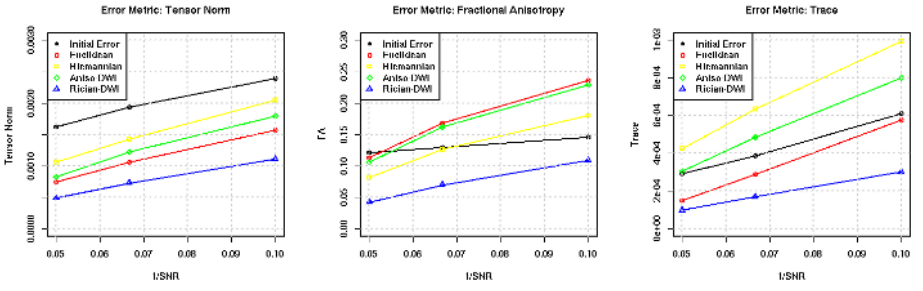


Fig. 3. Plots of error metrics for the various filters on synthetic data

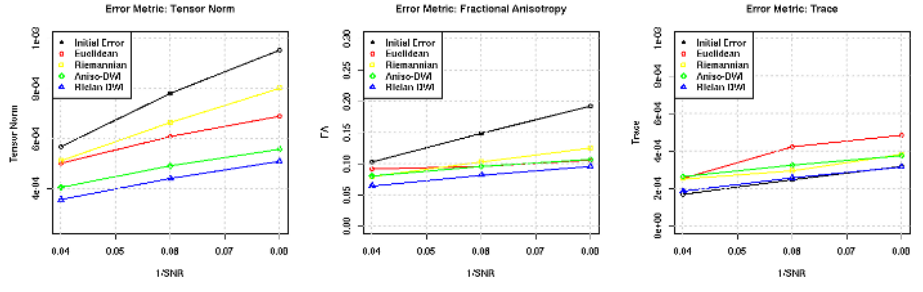


Fig. 4. Plots of error metrics for the various filters on real data

4.2 Real Data

DTI Ground Truth Generation: A key challenge in quantitatively evaluating filtering methods on real diffusion tensor data is the lack of a ground truth. While realistic simulated brain datasets exists for structural MRI, no such normalized data is available for DT-MRI. For this paper we develop a new approach for generating high SNR diffusion weighted image data. This technique builds a ground truth image as a maximum likelihood (ML) estimate from a set of repeated scans of the same subject. If $\{x_i\}$ is the set of intensities from the same voxel in N repeated scans, then the ML estimate of the true intensity A is found by maximizing the log-likelihood, $\log L = \sum_{i=1}^{N} \log p(x_i|A)$, where $p(x_i|A)$ is the Rician pdf as given in (1). The properties of the ML estimate are investigated by Sijbers *et al.* [11]. This ML estimate is superior to a simple averaging of the intensities as it incorporates *a priori* knowledge of the noise statistics. Also, it is well known that in the limit the ML estimate is most precise.

About the data: We generated our ground truth ML images from a set of five scans of a healthy volunteer on a Siemens head-only 3T scanner (Allegra). For each sequence, a single shot echo planar (EPI) diffusion tensor sequence with total scan time of approximately 12 minutes was used. The imaging parameters were: TR=5400ms, TR=73ms, isotropic voxels with 2mm slice distance and in-plane resolution = 2×2mm, 20 averages. We added known Rician noise to the

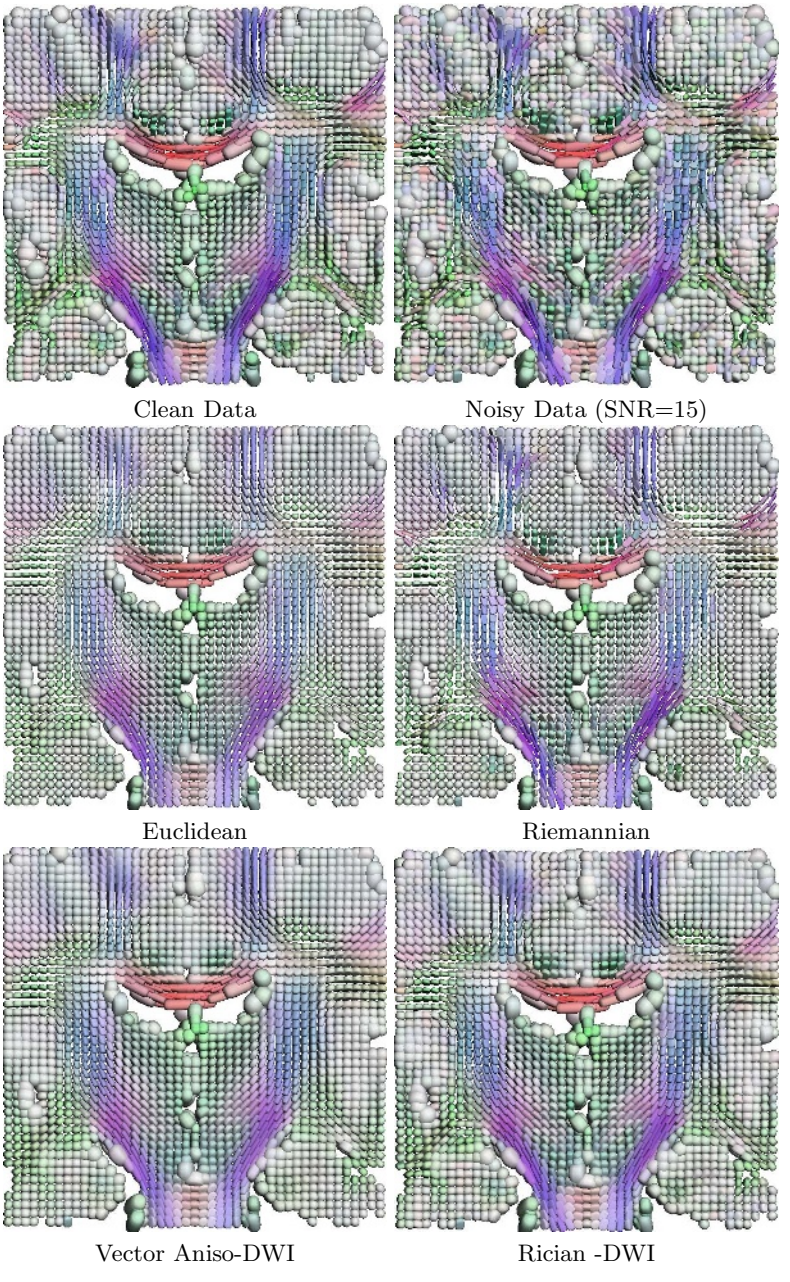


Fig. 5. Real Data Filtering Results

ML estimated DWIs at SNR levels of 10, 15 and 20 with respect to the average white matter signal level in the $b = 0$ image. A slice from the ground truth and SNR=15 tensor images is shown in Figure 5.

4.3 Performance

The resulting error metrics for the various filtering methods on the synthetic and real data are shown in Figures 3 and 4. The original, noisy and filtered images for SNR=15 are shown with superquadric glyphs [12] in Figures 2 and 5. The results demonstrate that the Rician filter with the bias correction term gives better RMS error on tensor components. On both the real and synthetic data the Rician filter is superior to all the other filtering techniques.The data also shows that for most of the error metrics the filtering methods on the DWI's yields better results than smoothing on the tensor space. The Riemannian filter requires all tensors to be positive definite and is thus disadvantaged by the process of adjusting for negative eigenvalues.

5 Conclusions

We presented a new method for denoising diffusion tensor images that includes a Rician noise model as part of MAP estimation framework. To the best of our knowledge, this is the first work to explicitly model and remove the bias effects of Rician noise in DT-MRI. We presented Monte Carlo simulations that show that noise can distort tensors in a manner that is dependent on the orientation and anisotropy of the underlying tensor. Our filtering results demonstrated that filtering on the original DWIs yields superior results to filtering methods that operate on the estimated tensors. Filtering on the DWIs with our Rician noise model gave the best overall results.

Acknowldgments

This work is part of the National Alliance for Medical Image Computing (NAMIC), funded by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant U54 EB005149. Information on the National Centers for Biomedical Computing can be obtained from `http://nihroadmap.nih.gov/bioinformatics`. Funding for this work has also been provided by Center for Integrative Biomedical Computing, NIH NCRR Project 2-P41-RR12553-07. We thank Weili Lin and Guido Gerig from the University of North Carolina for providing us with the DW-MRI data. Glyph visualizations created with Teem (`http://teem.sf.net`).

References

1. Pierpaoli, C., Basser, P.: Toward a quantitative assessment of diffusion anisotropy. Magnetic Resonance in Medicine **36**(6) (1996) 893–906
2. Anderson, A.W.: Theoretical analysis of the effects of noise on diffusion tensor imaging. Magnetic Resonance in Medicine **46**(6) (2001) 1174–1188
3. Skare, S., Li, T., Nordell, B., Ingvar, M.: Noise considerations in the determination of diffusion tensor anisotropy. Magnetic Resonance Imaging **18**(6) (2000) 659–669

4. J., P.G.: Nonlinear smoothing for reduction of systematic and random errors in diffusion tensor imaging. J Magn. Reson Imaging **11**(6) (2000) 702–710
5. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on Pattern Analysis Machine Intelligence **17**(4) (1990) 629–639
6. Wang, Z., Vemuri, B., Chen, Y., Mareci, T.: A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from complex DWI. IEEE Transactions on Medical Imaging **23**(8) (2004) 930–939
7. Martin Fernandez, C.F. Westin, C.A.L.: 3d bayesian regularization of diffusion tensor mri using multivariate gaussian markov random fields. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI'04). (2004)
8. Pennec, X., Fillard, P., Ayache, N.: A riemannian framework for tensor computing. International Journal of Computer Vision **66**(1) (2006) 41–66
9. Fillard, P., Arsigny, V., Pennec, X., Ayache, N.: Clinical DT-MRI estimation, smoothing and fiber tracking with log-Euclidean metrics. In: Proceedings of the Third IEEE International Symposium on Biomedical Imaging (ISBI 2006), Crystal Gateway Marriott, Arlington, Virginia, USA (2006) 786–789
10. Basser, P., Pierpaoli, C.: Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. J. Mag Res. **111**(3) (1996) 209–219
11. Sijbers, J., den Dekker, A., Scheunders, P., Dyck, D.V.: Maximum-Likelihood Estimation of Rician Distribution Parameters. IEEE Transactions on Medical Imaging **17**(3) (1998) 357–361
12. Kindlmann, G.: Superquadric tensor glyphs. In: Proceedings of IEEE TVCG/EG Symposium on Visualization 2004. (2004) 147–154

Anisotropy Creases Delineate White Matter Structure in Diffusion Tensor MRI

Gordon Kindlmann[1], Xavier Tricoche[2], and Carl-Fredrik Westin[1]

[1] Laboratory of Mathematics in Imaging, Department of Radiology, Harvard Medical School, USA
gk@bwh.harvard.edu
[2] Scientific Computing and Imaging Institute, University of Utah, USA

Abstract. Current methods for extracting models of white matter architecture from diffusion tensor MRI are generally based on fiber tractography. For some purposes a compelling alternative may be found in analyzing the first and second derivatives of diffusion anisotropy. *Anisotropy creases* are ridges and valleys of locally extremal anisotropy, where the gradient of anisotropy is orthogonal to one or more eigenvectors of its Hessian. We propose that anisotropy creases provide a basis for extracting a skeleton of white matter pathways, in that ridges of anisotropy coincide with interiors of fiber tracts, and valleys of anisotropy coincide with the interfaces between adjacent but distinctly oriented tracts. We describe a crease extraction algorithm that generates high-quality polygonal models of crease surfaces, then demonstrate the method on a measured diffusion tensor dataset, and visualize the result in combination with tractography to confirm its anatomic relevance.

1 Introduction

Diffusion tensor magnetic resonance imaging (DTI) is a popular means of assessing white matter in the central nervous system. Coherent organization of axons leads to diffusion anisotropy, and insofar as a tensor model accurately represents the form and direction of anisotropy, DTI can detect white matter architecture [1]. Fiber tractography has become the dominant method of DTI analysis, wherein the course of axons in fiber tracts is modeled by computing paths along the direction of greatest diffusivity (the diffusion tensor principal eigenvector), allowing connectivity between different brain regions to be mapped and quantified [2,3]. Clustering collects individual tracts into coherent structures that can model the shape and direction of fiber pathways [4]. Other white matter analysis methods do not use connectivity information from tractography, e.g. region-of-interest studies of fractional anisotropy (FA) [5,6].

We introduce *anisotropy creases* as a technique for extracting a skeleton of white matter directly from the intrinsic structure of FA. Scalar image processing defines *creases* as features at which the gradient is orthogonal to one or more eigenvectors of the Hessian [7]. Given the ubiquity of FA as a quantitative variable in the diffusion tensor literature, we have started by detecting creases

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 126–133, 2006.

in FA. We propose that the ridge surfaces and ridge lines of FA coincide with the interiors of white matter fiber tracts, and that valley surfaces of anisotropy delineate the interfaces between fiber tracts that are adjacent but orthogonally oriented (such as between the corpus callosum and the cingulum bundles).

Anisotropy creases may have utility in a variety of contexts. Explicitly modeling the interfaces between adjacent but orthogonal fiber tracts may helpfully constrain non-rigid registration of tensor fields for group studies, as slight misregistration of these configurations could lead to comparison of tensor values within entirely separate pathways. The ability to extract white matter skeletons directly from tensor invariants, without the algorithmic complexity or parameter tuning of fiber tracking and clustering, could increase sensitivity in shape analysis studies. Finally, major crease features could play a role analogous to that of the cortical surface in functional imaging, namely a reference manifold onto which variables of interest are projected and analyzed [8].

2 Related Work

Creases have been an object of study for many years in different disciplines. In the context of geomorphology, de Saint-Venant [9] defines creases as the loci where the slope is minimal along the isocontours of the relief, which Haralick later reformulates in terms of the Hessian of the height function [10]. Maxwell gives a topological and global definition of ridges and valleys as watersheds and watercourses: slope lines that connect saddle points to local maxima or minima [11]. Others study creases in terms of differential geometry [12]. Eberly *et al.* motivate the idea that creases should be defined locally and be invariant with respect to a variety of transforms (rigid transforms, uniform scaling, and monotonic mappings of intensity) [7]. They also generalize the height-based definition of de Saint-Venant to d-dimensional manifolds embedded in n-dimensional image space, and observe that this definition produces good results for a medical imaging problem [7]. Other previous work focuses on extracting polygonal models of crease geometry; this is reviewed in Section 3.3.

A separate line of previous work studies feature detection in DTI by means other than tractography. Pajevic *et al.* use B-splines to generate continuous tensor fields that are differentiated to highlight anisotropy boundaries [13]. O'Donnell *et al.* use structure tensors to detect general boundaries in tensor values [14]. In both cases, results are visually evaluated by confirming a high edge strength near structural boundaries, but the techniques do not analyze the familiar FA measure, nor is the feature geometry explicitly extracted. Recent work by Smith *et al.* is most similar to our approach in that they perform voxel-based morphometry with a white matter skeleton calculated from ridges in a mean FA map (though "ridges" are not mentioned per se) [8]. By using an established mathematical definition of crease features, our technique extracts true codimension-one crease surfaces from continuous tensor fields of individual DTI scans, rather than a voxel-based ridge mask from the inherently smooth mean FA image from a set of registered scans.

3 Methods

3.1 Convolution and Differentiation

We start with a first-principles approach to measuring derivatives of FA in a tensor field. We use separable convolution between C^2 cubic splines and the tensor coefficient discrete samples to reconstruct a smooth tensor field [15,13]. By linearity, analytic derivatives of the reconstructed field are measured by convolving the sampled data with derivatives of the reconstruction kernel [16]. FA can be defined in terms of principal invariants J_i, which can in turn be expressed in terms of the tensor coefficients [5]. Differentiating these relations produces formulae for the gradient of FA in terms of the gradients of tensor coefficients:

$$\mathrm{FA} = \sqrt{\frac{J_4 - J_2}{J_4}} \; ; \; \begin{array}{l} J_2 = D_{xx}D_{yy} + D_{xx}D_{zz} + D_{yy}D_{zz} - D_{xy}^2 - D_{xz}^2 - D_{yz}^2 \\ J_4 = D_{xx}^2 + D_{yy}^2 + D_{zz}^2 + 2D_{xy}^2 + 2D_{xz}^2 + 2D_{yz}^2 \end{array} \tag{1}$$

$$\nabla\mathrm{FA} = \frac{J_2 \nabla J_4 - J_4 \nabla J_2}{2J_4^2\sqrt{1 - \frac{J_2}{J_4}}} \; ; \; \begin{array}{rl} \nabla J_2 = & (D_{yy}+D_{zz})\nabla D_{xx} + (D_{xx}+D_{zz})\nabla D_{yy} + (D_{xx}+D_{yy})\nabla D_{zz} \\ & - 2D_{xy}\nabla D_{xy} - 2D_{xz}\nabla D_{xz} - 2D_{yz}\nabla D_{yz} \\ \nabla J_4 = & 2D_{xx}\nabla D_{xx} + 2D_{yy}\nabla D_{yy} + 2D_{zz}\nabla D_{zz} \\ & + 4D_{xy}\nabla D_{xy} + 4D_{xz}\nabla D_{xz} + 4D_{yz}\nabla D_{yz} \end{array} \tag{2}$$

Formulae for the second derivative (the Hessian) are longer but straight-forward to derive. Because FA is a non-linear function of the tensor, pre-computing FA on a discrete grid and then differentiating is *not* equivalent to this approach.

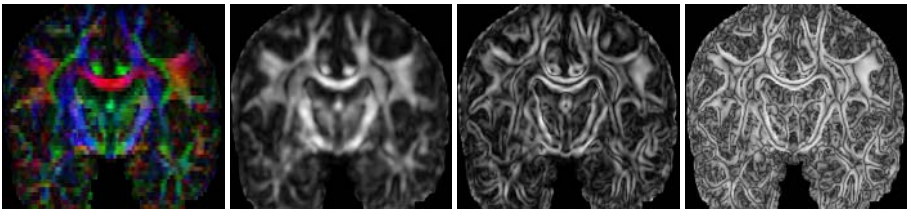
3.2 Crease Feature Definition

Crease features are defined in terms of the gradient $\mathbf{g} = \nabla f$ and Hessian $\mathbf{H}$ of a scalar field f [7]. Section 3.1 described how to measure the derivatives of FA at any point in a tensor field. Let $\lambda_1 \geq \lambda_2 \geq \lambda_3$ be the eigenvalues of $\mathbf{H}$, and $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ be the corresponding eigenvectors. Ridges and valleys are defined by:

	Surface	Line
Ridge	$\mathbf{g}\cdot\mathbf{e}_3 = 0,\ \lambda_3 < 0$	$\mathbf{g}\cdot\mathbf{e}_2 = \mathbf{g}\cdot\mathbf{e}_3 = 0,\ \ \lambda_3, \lambda_2 < 0$
Valley	$\mathbf{g}\cdot\mathbf{e}_1 = 0,\ \lambda_1 > 0$	$\mathbf{g}\cdot\mathbf{e}_1 = \mathbf{g}\cdot\mathbf{e}_2 = 0,\ \ \lambda_1, \lambda_2 > 0$

One way to inspect ridge surfaces (for example) in a volume is to densely sample $|\mathbf{g}\cdot\mathbf{e}_3|$ on a two-dimensional cutting plane and look for the dark lines indicating the ridge surface intersection with the plane. Figure 1 shows an example of FA ridge surfaces in a two-dimensional coronal slice of a human brain DTI scan[1]. For context, Figure 1(a) shows the standard RGB colormap of $\mathbf{e}_1$ at the original image resolution. Figures 1(b) and 1(c) illustrate how smooth features arise from convolution-based measurements of FA and ∇FA. Figure 1(d) uses contrast-enhancement to show the dark smooth lines indicating the ridge surfaces. The cubic spline used in this and all other results was 4mm between inflection points.

[1] DTI data was estimated from 30 DWIs at b = 700s/mm^2 and 5 non-DWI T2s, from a 1.5 T Philips scanner, with resolution $0.94 \times 0.94 \times 2.5$mm.



(a) RGB($\mathbf{e}_1$) (b) FA (c) $\mathbf{g} = |\nabla(\text{FA})|$ (d) $|\mathbf{g} \cdot \mathbf{e}_3|^{1/3}$

Fig. 1. Two-Dimensional Demonstration of Ridge Surface Evaluation

3.3 Crease Surface Extraction

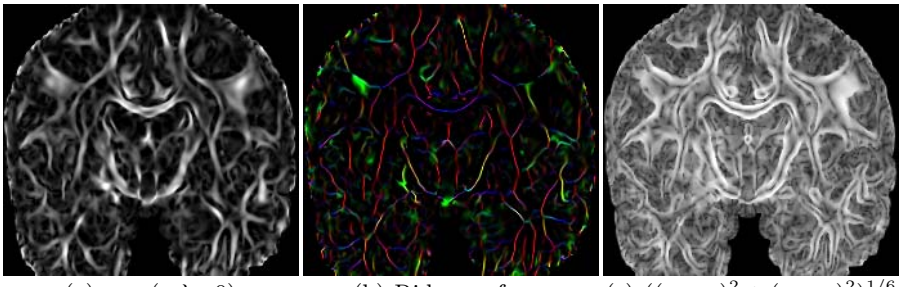
We extract crease surfaces by per-voxel triangulation of the zero-isocontour of $\mathbf{g} \cdot \mathbf{e}_i$ ($\mathbf{g} \cdot \mathbf{e}_3$ for ridges, $\mathbf{g} \cdot \mathbf{e}_1$ for valleys) using Marching Cubes [17], taking care in evaluating $\mathbf{g} \cdot \mathbf{e}_i$ at voxel corners. Eigenvectors lack intrinsic sign, so $\mathbf{g} \cdot \mathbf{e}_i$ can suddenly change sign, far from a crease, simply due to the numerical properties of eigenvector computation. The literature offers ways to overcome this. Morse suggests determining correspondences between sets of eigenvectors rather than individual ones, to handle eigenvector permutations associated with eigenvalue equality [18]. Furst *et al.* use similar ideas in *Marching Cores* to extract crease manifolds in image scale-space [19]. For *Marching Ridges*, Furst and Pizer choose eigenvector signs to agree with the principal eigenvector of the average of outer products of the eigenvectors under consideration [20].

Our experience suggests that Hessian eigenvectors of non-linear scalar attributes of tensors (such as FA) tend to vary more rapidly than those of a similarly sampled scalar field. Accordingly, we explicitly determine eigenvector orientation consistency by traversing voxel edges to monitor eigenvector rotation. Samples are adaptively generated along voxel edges to satisfy an upper-bound on the angle between unsigned eigenvectors at successive samples. This determines whether the smooth transport of $\mathbf{e}_i(\mathbf{v}_0)$ from vertex $\mathbf{v}_0$ to vertex $\mathbf{v}_1$ agrees in sign with the eigenvector $\mathbf{e}_i(\mathbf{v}_1)$ computed at $\mathbf{v}_1$. The per-edge eigenvector sign information determines a per-voxel sign consistency prior to evaluating the Marching Cubes case. Inter-voxel sign inconsistencies lead to triangulations with inconsistent vertex windings. Thus, as a final pass, we traverse the surface mesh to fix vertex windings, which allows graphics hardware to appropriately render the crease surfaces with two-sided lighting [21].

The continuous tensor field measurements allow the voxel grid of the crease surface triangulation to be independent of the underlying data resolution. The results in Section 4 use a triangulation resolution two to three times that of the data. The strength of the crease surface is assessed with the appropriate Hessian eigenvalue ($-\lambda_3$ for ridges, λ_1 for valleys) so that geometry is extracted only for significant features.

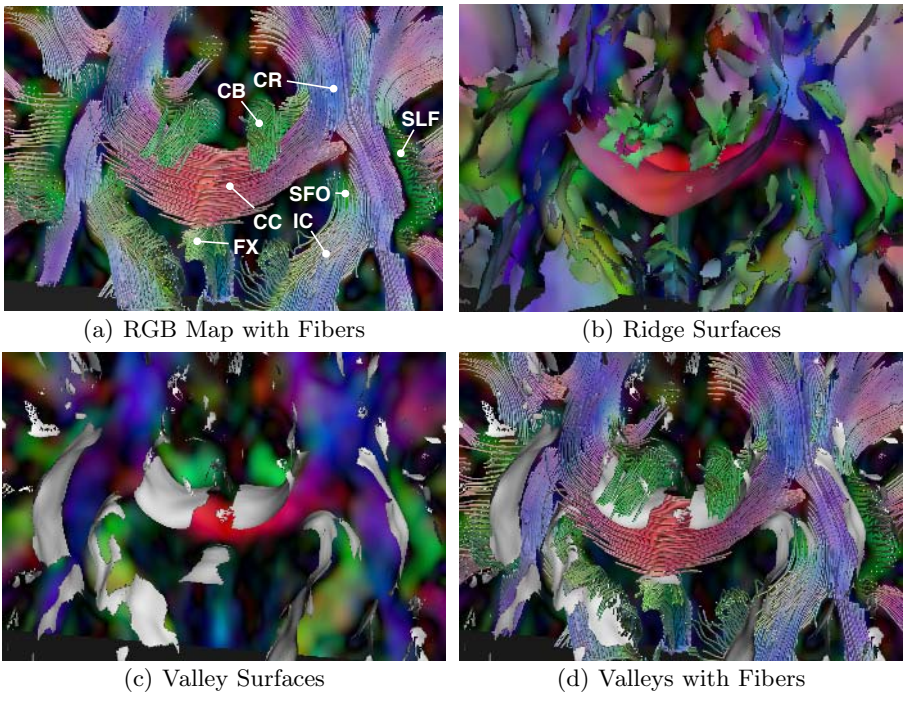
4 Results

Figure 2 shows ridge detection results on the same cutting plane used in Figure 1. Ridge strength is mapped in Figure 2(a), the coherent organization of which is



(a) $\max(-\lambda_3, 0)$ (b) Ridge surfaces (c) $((\mathbf{g} \cdot \mathbf{e}_2)^2 + (\mathbf{g} \cdot \mathbf{e}_3)^2)^{1/6}$

Fig. 2. Two-Dimensional Ridge Surface and Ridge Line Results



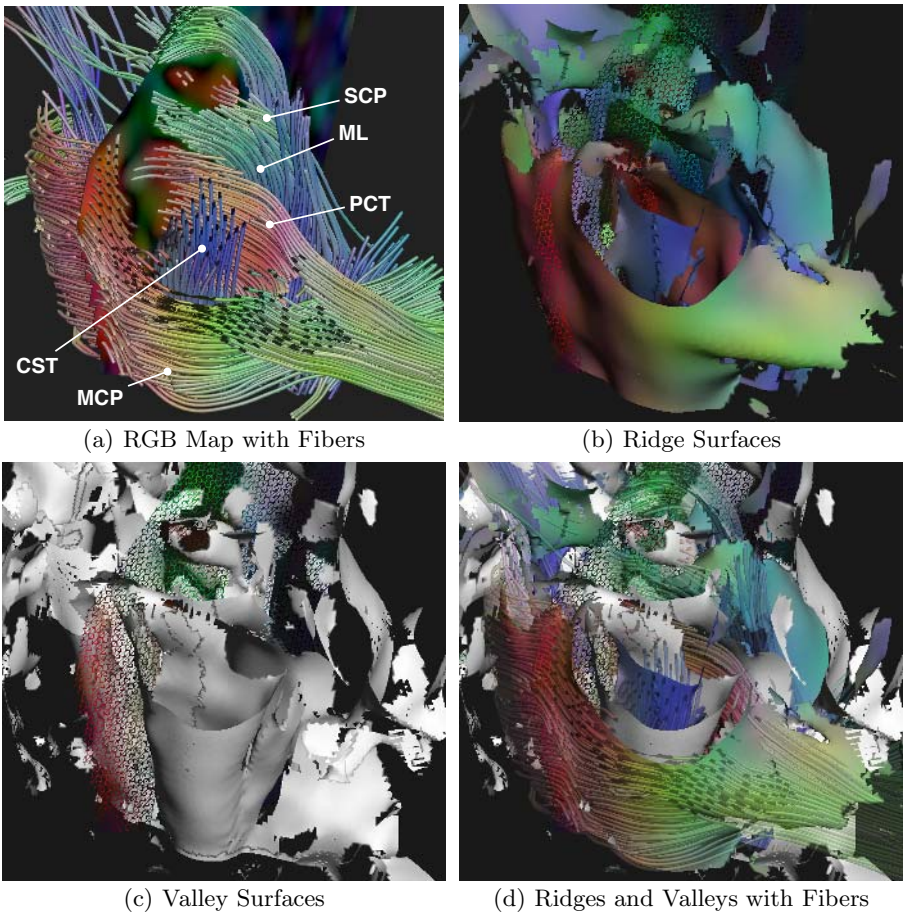
(a) RGB Map with Fibers (b) Ridge Surfaces

(c) Valley Surfaces (d) Valleys with Fibers

Fig. 3. Anisotropy Creases Near the Corpus Callosum

suggestive of white matter pathways. A combination of ridge strength masking, contrast enhancement, and RGB directional encoding created Figure 2(b) from Figure 1(d), to highlight the lines through major white matter pathways. Figure 2(c) maps a quantity that is zero on ridge *lines* of FA, intersecting the plane in points. Multiple lines lie within the corpus callosum, but there are two clear dots within the cingulum bundles, and one in the fornix, consistent with the shape and orientation of these tracts relative to this coronal plane.

The renderings in Figure 3 (from a posterior viewpoint) show a cropped region around the same coronal slice of previous figures. In Figure 3(a) fibers are seeded



(a) RGB Map with Fibers (b) Ridge Surfaces

(c) Valley Surfaces (d) Ridges and Valleys with Fibers

Fig. 4. Anisotropy Creases in the Brainstem

from the RGB encoded plane. Figure 3(b) shows how the ridge surfaces (using the same RGB encoding) follow major fiber paths, especially the corpus callosum (CC), internal capsule (IC), corona radiata (CR), and fornix (FX) [22]. The (white) anisotropy valley surfaces in Figure 3(c) delineate interfaces between the CC and cingulum bundles (CB), superior fronto-occipital fasciculus (SFO) and IC, and IC and superior longitudinal fasciculus (SLF). Figure 3(d) also illustrates how anisotropy valleys lie between adjacent paths of differing orientation.

Figure 4 illustrates anisotropy crease analysis in the brainstem (lateral anterior superior viewpoint), starting with (in Figure 4(a)) a plane seeding fibers for anatomical context. Here, a viewpoint-aligned cutting plane partially cuts into the middle cerebellar peduncle (MCP) to reveal the corticospinal tract (CST) and pontine crossing tract (PCT), anterior to the medial lemniscus (ML) and superior cerebellar peduncle (SCP). These pathways appear as distinct anisotropy ridge surfaces in Figure 4(b), and their interfaces are delineated by the valley

surfaces in Figure 4(c), especially the enclosure of CST between MCP and PCT. Figure 4(d) combines the crease surfaces with faint fibers to illustrate how the alternating layers of ridges and creases combine to form a fiber path skeleton.

5 Discussion and Future Work

Anisotropy creases model white matter structure from DTI using continuously measured anisotropy derivatives and explicitly triangulated surface geometry. The invariance properties of their mathematical definition help give anisotropy creases the attractive property of being parameter free (aside from the choice of convolution kernels), in contrast to most tractography and clustering algorithms.

Future work will include extraction of FA ridge *lines*, and their comparison to individual fiber tracts. An interesting question left unanswered is why strong valley surfaces of FA reliably indicate adjacent orthogonal tracts, considering that FA is simply a scalar invariant, with no knowledge of nearby eigenvectors. Finally, to extract true image *cores*, crease detection must work across measurement scales [19], which we have not yet implemented. Prior anatomical knowledge may usefully constrain, however, the necessary scale range. For example, extracting fiber interfaces as anisotropy valleys may require only a single image scale, considering that the interfaces are not apt to have much physical thickness, thus the measurement scale is determined by the acquisition resolution.

We hope to use anisotropy creases with non-rigid tensor registration, first analyzing existing methods according to how well anisotropy creases are aligned, and then perhaps enhancing registration to use the creases as fiducials.

Acknowledgements

This work supported by NIH NIBIB T32-EB002177, NIH NCRR P41-RR13218 (NAC) and 2-P41-RR12553-07 (CIBC), and NIH R01-MH050740. DWI data courtesy of Dr. Susumu Mori, Johns Hopkins University, supported by NIH R01-AG20012-01 and P41-RR15241-01A1.

References

1. Basser, P., Jones, D.: Diffusion-tensor MRI: Theory, experimental design and data analysis - A technical review. Nuc Mag Res in Biomed **15** (2002) 456–467
2. Mori, S., Zijl, P.V.: Fiber tracking: Principles and strategies - A technical review. Nuc Mag Res in Biomed **15** (2002) 468–480
3. Corouge, I., Fletcher, P., Joshi, S., Gilmore, J., Gerig, G.: Fiber tract-oriented statistics for quantitative diffusion tensor MRI analysis. In: Eighth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'05). (2005) 131–139
4. Moberts, B., Vilanova, A., van Wijk, J.: Evaluation of fiber clustering methods for diffusion tensor imaging. In: Proceedings of IEEE Visualization 2005. (2005) 65–72

5. Basser, P.: Inferring microstructural features and the physiological state of tissues from diffusion-weighted images. Nuc Mag Res in Biomed **8** (1995) 333–344
6. Kubicki, M., Westin, C.F., Maier, S., Mamata, H., Frumin, M., Ernst-Hirshefeld, H., Kikinis, R., Jolesz, F., McCarley, R., Shenton, M.: Cingulate fasciculus integrity disruption in schizophrenia: A magnetic resonance diffusion tensor imaging study. Biological Psychiatry **54** (2003) 1171–1180
7. Eberly, D., Gardner, R., Morse, B., Pizer, S.: Ridges for image analysis. Journal of Mathematical Imaging and Vision **4** (1994) 351–371
8. Smith, S., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T., Mackay, C., Watkins, K., Ciccarelli, O., Cader, M., Matthews, P., Behrens, T.: Tract-based spatial statistics: Voxelwise analysis of multiple-subject diffusion data. Technical Report TR05SS1, Oxford University Centre for Functional MRI of the Brain (2006)
9. de Saint-Venant, M.: Surfaces à plus grande pente constituées sur des lignes courbes. Bulletin de la Société Philomathématique de Paris (1852) 24–30
10. Haralick, R.M.: Ridges and valleys on digital images. Computer Vision, Graphics, and Image Processing **22** (1983) 28–38
11. Maxwell, J.: On hills and dales. The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science **40**(269) (1870) 421–425
12. Gauch, J.M., Pizer, S.M.: Multiresolution analysis of ridges and valleys in grey-scale images. IEEE Trans Pat Analysis and Mach Int **15**(6) (1993) 635–646
13. Pajevic, S., Aldroubi, A., Basser, P.: A continuous tensor field approximation of discrete DT-MRI data for extracting microstructural and architectural features of tissue. Journal of Magnetic Resonance **154** (2002) 85–100
14. O'Donnell, L., Grimson, W., Westin, C.F.: Interface detection in diffusion tensor MRI. In: Seventh International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'04). (2004) 360–367
15. Aldroubi, A., Basser, P.: Reconstruction of vector and tensor fields from sampled discrete data. Contemporary Mathematics **247** (1999) 1–15
16. Gonzalez, R., Woods, R.: Digital Image Processing. 2nd edn. Addison-Wesley Publishing Company, Reading, MA (2002)
17. Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3D surface construction algorithm. Computer Graphics **21**(4) (1987) 163–169
18. Morse, B.S.: Computation of Object Cores from Grey-Level Images. PhD thesis, University of North Carolina at Chapel Hill, Chapel Hill, NC (1994)
19. Furst, J.D., Pizer, S.M., Eberly, D.H.: Marching cores: A method for extracting cores from 3d medical images. In: Proceedings of IEEE Workshop on Mathematical Methods in Biomedical Image Analysis. (1996) 124–130
20. Furst, J.D., Pizer, S.M.: Marching ridges. In: Proceedings of 2001 IASTED International Conference on Signal and Image Processing. (2001)
21. Shreiner, D., Woo, M., Neider, J., Davis, T.: OpenGL Programming Guide. Fourth edn. Addison-Wesley (2004)
22. Mori, S., Wakana, S., Nagae-Poetscher, L., Zijl, P.V.: MRI Atlas of Human White Matter. Elsevier (2005)

Evaluation of 3-D Shape Reconstruction of Retinal Fundus

Tae Eun Choe[1], Isaac Cohen[1], Gerard Medioni[1],
Alexander C. Walsh[2], and SriniVas R. Sadda[2]

[1] IRIS, University of Southern California
3737 Watt way, Los Angeles, CA, 90248, USA
{tchoe, icohen, medioni}@usc.edu
[2] Doheny Eye Institute,
1450 San Pablo Street, Los Angeles, CA, 90033, USA
{AWalsh, SSadda}@doheny.org

Abstract. We present a method for the 3-D shape reconstruction of the retinal fundus from stereo paired images. Detection of retinal elevation plays a critical role in the diagnosis and management of many retinal diseases. However, since the shape of ocular fundus is nearly planar, its 3-D depth range is very narrow. Therefore, we use the location of vascular bifurcations and a plane+parallax approach to provide a robust estimation of the epipolar geometry. Matching is then performed using a mutual information algorithm for accurate estimation of the disparity maps. To validate our results, in the absence of camera calibration, we compared the results with measurements from the current clinical gold standard, optical coherence tomography (OCT).

1 Introduction

Retinal diseases are the most common causes of acquired blindness among adults in developed nations. Glaucoma is also a frequent cause of vision loss worldwide. One characteristic feature of all of these diseases is their distortion, either through excavation, elevation or thickening, of the ocular fundus topography. For this reason, 3-D visualization or imaging of the retina and optic nerve are necessary for the accurate diagnosis and evaluation of these disorders. Traditionally, this assessment has been performed by highly-trained eye-care professionals using examining microscopes. More recently, an imaging technique known as optical coherence tomography (OCT) has emerged for this purpose. Based on the principle of low-coherence interferometry, OCT provides an *in vivo* cross-sectional image of the retina that simulates microscopic visualization and has axial resolutions under 3μm [11]. Unfortunately, the cost of OCT equipment and the expertise required for its interpretation has limited widespread adoption of this technology.

An alternative to subjective inspection and high-cost imaging consists of inferring the 3D shape of the retina using images acquired with a common, lower-cost fundus camera. This paradigm suggests a potentially cost-effective solution for developing nations faced with the emerging epidemic of diabetes mellitus. Since this analysis can

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 134 – 141, 2006.

be carried out either on-site or remotely, it is also well-suited to a modern era of global telemedicine.

In the assessment of pathologic changes due to retinal disease or glaucoma, the 3-D reconstruction of two main areas of the ocular fundus is important: the optic disc and the macula. The numerous features provided by blood vessels merging at the optic disc make this area easier to reconstruct than the macula which is equally important in human vision. For example, the normal macula is devoid of large vessels and other consistent registration points. Therefore, stereoscopic reconstructions of this essential part of the eye are quite challenging. Retinal images also have unique features that prevent classical stereo algorithms from estimating the 3-D shape of the fundus from stereo paired images. Intensities and colors of the same physical positions may vary between consecutive images. In addition, the shape of ocular fundus is nearly planar and so, its 3-D depth is very narrow. This prevents accurate estimation of the epipolar geometry using traditional stereo algorithms. Finally, image blur may prevent accurate stereo reconstructions.

Scharstein and Szeliski present a comprehensive review of commonly used methods for stereo estimation [12]. Graph cut [7] and belief propagation [13] methods have shown good performance on test sets of stereo images which deal with scenes with large depth changes, similar intensities for identical pixels, and textured regions.

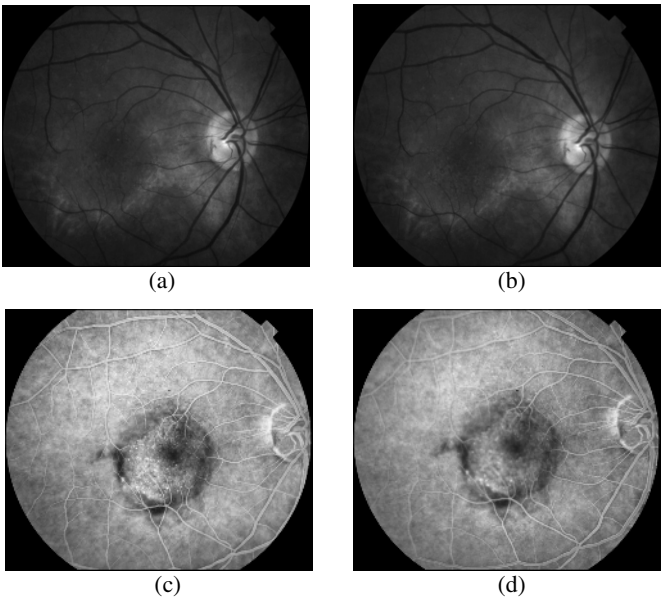


Fig. 1. Two pairs of stereo retinal images from a patient with age-related macular degeneration (a)(b) A characteristic 'blister' in the macula is harder to see in the red-free images (a)(b) than it is in the images taken after injection of fluorescein dye (c)(d) into the antecubital vein. Nevertheless, the image in (d) is blurred which complicates stereoscopic measurements.

As discussed above, retinal images have challenging features such as inconsistent intensities, smooth and thin surfaces, and less texture in the image, that complicate stereoscopic estimations. Methods commonly used in the literature are not well adapted for stereo estimation of the retina from fundus images, as attested by our comparative study presented in the experimental results section. Figure 1 shows various sets of stereo images.

Our paper proposes a method for 3-D reconstruction from a pair of retinal images and its evaluation. The method consists of three steps. First, using matched Y-features, a plane+parallax approach is taken to estimate the epipolar geometry. After image rectification, the search space on the scanline for stereo matching is estimated based on the Y-feature correspondences. Second, our mutual information method matches points to produce a dense disparity map. Finally, the reconstructed 3-D shape is validated by OCT data.

2 Estimation of Epipolar Geometry

This section briefly describes the method for estimating the epipolar geometry from a pair of fluorescein images of the retina [1]. Many types of robust features can be used to estimate the epipolar geometry such as Harris corners [4] or SIFT features [10]. We utilized the locations of vessels' bifurcations, so-called Y-features, which are well distributed in the fundus and more robust to intensity changes within images. These features are used to estimate the epipolar geometry and search space in stereo matching. Extracted Y-features are matched by maximizing a mutual information criterion.

Typically, at least seven point correspondences are necessary to estimate the fundamental matrix of the epipolar geometry. However, various standard implementations of the 7-points and 8-points algorithm were tested and did not provide satisfactory results [9]. Figure 2-(a)(c) shows the erroneous result of the

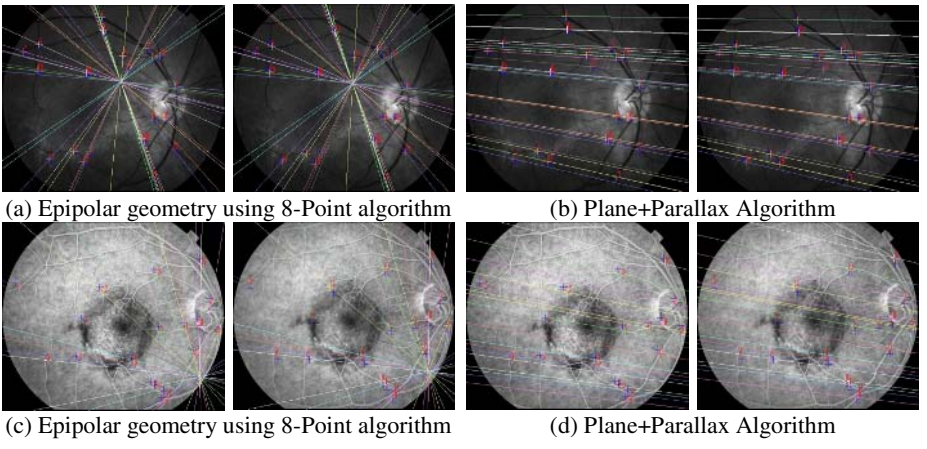


Fig. 2. (a)(c) Examples of inaccurately estimated epipolar geometry with 8-point algorithm in the case of a translation of the camera. (b)(d) The epipolar lines obtained from the plane+parallax based fundamental matrix estimation.

corresponding epipolar lines defined by the fundamental matrix estimated with an 8-point algorithm. In this example, the motion of the camera between the two acquisitions is close to a pure translation. However, a degenerate case occurs in the estimation of the fundamental matrix when all the points lie on a plane. This happens quite frequently in retinal imaging since the surface of the retina is relatively flat. Although the points are not on the same plane, the range of 3-D depth appears to be too narrow to get a satisfactory fundamental matrix from the 8-point algorithm.

To overcome this limitation of the 8-point algorithm on retinal images, we implemented the plane+parallax algorithm proposed by Kumar et al [8]. Given 4 corresponding points, first a homography is calculated. Adding 2 more point correspondences belonging to residual parallax regions enables us to estimate the location of the epipoles. We have implemented a RANSAC-based algorithm for the plane+parallax method. Fig. 2-(b)(d) shows the corresponding epipolar lines by the plane+parallax approach. After estimating the fundamental matrix, stereo images are rectified using Gluckman and Nayar's method [3] for the estimation of the depth map.

3 Stereo Matching

After rectification, all matching pixels are located on the same scanlines. On the scanline, the lower and the upper bound of the search space is estimated by utilizing disparities of the matched Y-features [2]. This method produces a 1-D narrow search space, which enables a more accurate computation of the dense disparity map. In the case of retinal imaging where the intensities of the matching areas vary across the stereo images, a general cross-correlation based algorithm did not seem to provide satisfactory results. Instead, we implemented a mutual information algorithm for point matching along the scanlines and estimation of the depth map.

Mutual information for a pair of windows $\mathbf{z}_A$ and $\mathbf{z}_B$ is defined by [14]:

$$MI(\mathbf{z}_A, \mathbf{z}_B) = H(\mathbf{z}_A) + H(\mathbf{z}_B) - H(\mathbf{z}_A, \mathbf{z}_B) \tag{1}$$

where $H(\mathbf{z}) = -\int p(\mathbf{z}) \ln p(\mathbf{z})\, d\mathbf{z}$ is Shannon entropy of the image window $\mathbf{z}$, and p is the distribution of the grey levels in the considered window. In [6], the authors used only joint entropy $H(\mathbf{z}_A, \mathbf{z}_B)$ rather than mutual information. The proposed method does not work well on low-textured areas. When mostly textureless areas are compared to each other, joint entropy has a high value on textureless area, which is incorrect. However marginal entropies $H(\mathbf{z}_A)$ and $H(\mathbf{z}_B)$ help to boost the mutual information value on the textured structures. We consider the following approximation of the entropy [14]:

$$H(\mathbf{z}) \approx -\frac{1}{N_{\mathbf{z}}} \sum_{z_i \in \mathbf{z}} \ln p(z_i)$$

where $N_{\mathbf{z}}$ is the size of the window $\mathbf{z}$, and the density function $p(z)$ is estimated based on Parzen window density estimation. We consider a Gaussian density function for the Parzen window W_P, and the distribution of the grey levels is locally approximated as follows:

$$p(z) \approx \frac{1}{N_p} \sum_{z_j \in W_P} g_\psi (z - z_j)$$

where N_p is the number of sample in the Parzen window W_P, and $g_\psi(z)$ is the uni- or bi-variate Gaussian density function with diagonal covariance matrix ψ [14]. In [5], the author estimated the discretized density function $p(z)$ in order to reduce the time complexity. $p(z)$ is calculated at the given point and it is convoluted by 2-D Gaussian to simulate the density function. However this simplification reduces the accuracy of the mutual information. The entropy function is then rewritten as:

$$H(\mathbf{z}) \approx -\frac{1}{N_\mathbf{z}} \sum_{z_i \in \mathbf{z}} \ln \frac{1}{N_p} \sum_{z_j \in W_P} g_\psi (z_i - z_j) \tag{2}$$

We utilize Equation (2) to implement the mutual information. In mutual information, the marginal entropies $H(\mathbf{z}_A)$ and $H(\mathbf{z}_B)$ are included and the density function *p(z)* is calculated individually using Parzen window estimation with Gaussian assumption. The disparity of each pixel is determined by following equation.

$$disparity = \arg\max_{d \in S} MI(d) = \arg\max_{d \in S} MI(\mathbf{z}_A, \mathbf{z}_{B,d}) \tag{3}$$

where $MI(\mathbf{z}_A, \mathbf{z}_{B,d})$ is the mutual information from Equation (1) and $\mathbf{z}_{B,d}$ is the window of d pixel distance from the window $\mathbf{z}_B$ in the second image. Each pixel is centered in a window $\mathbf{z}_A$, which then is compared with a window $\mathbf{z}_{B,d}$ in the other image within the search space S. The pair of windows that has the maximum mutual information determines the disparity of each pixel position.

To speed up the calculation of MI, we constructed a look-up table of Gaussian function in advance. Since the heaviest calculation of MI is the Gaussian function and the variable is an integer that ranges from 0 to 255, for each input value, the Gaussian function was pre-calculated and saved in the look-up table. The same method was applied to the 2-D Gaussian function, which requires two integer input variables. This implementation made the calculation time of MI 8 times faster than the regular MI implementation.

Since the range of disparity values is too narrow, the subpixel resolution of disparity is essential in 3-D reconstruction of retinal images to avoid "staircasing" in the disparity map. We estimate the disparity map using subpixel accuracy by using a quadric interpolation relying on neighboring mutual information values. Every disparity is determined individually by mutual information matching criteria. We did not apply a smoothness constraint to calculate disparities. Based on the taxonomy of stereo algorithms [12], only the matching function is applied. Neither aggregation nor optimization methods are used. In our experiments we have noted that these methods degraded the accuracy of the matching.

4 Evaluation of 3-D Surface

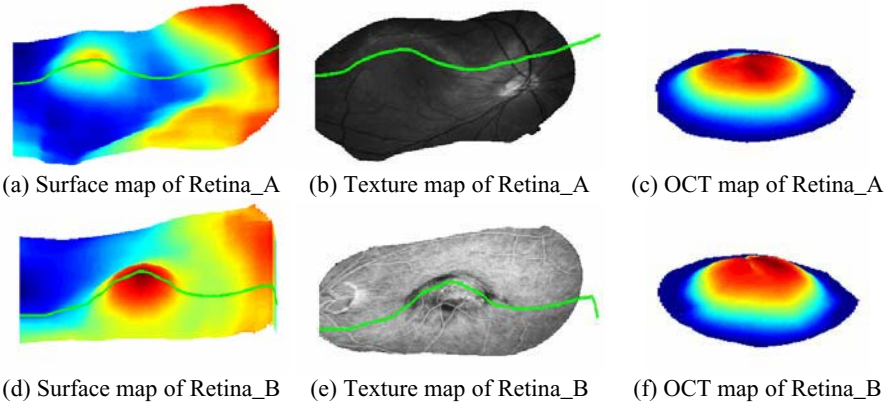
The reconstructed 3-D shape is evaluated by comparing our results to aligned OCT data from a commercial instrument (Carl Zeiss Meditec, Dublin, CA). Our data differs slightly from OCT data since we have a disparity map – not a true depth map. A more

accurate depth map requires knowledge of internal camera parameters that are currently unavailable.

4.1 3-D Shape of Retinal Fundus

We conducted experiments on numerous stereo pairs and selected three pairs, named Retina_A, B and C, for evaluation. Fig. 1 presents the two selected pairs of stereo images that had corresponding OCT data. The images demonstrate a retinal pigment epithelial detachment ('blister') in the macula secondary to choroidal neovascularization from age-related macular degeneration. The 'blister' in the right stereo image of Retina_B is unfocused, which makes the image reconstruction difficult.

After Y-features were extracted and matched from each image, the epipolar geometry was estimated with the plane+parallax algorithm (Fig. 2). The stereo images were rectified based on the fundamental matrix, and the search space was estimated from a set of matching Y-features. Using mutual information, subpixel resolution dense disparity maps were estimated (Fig. 3).



(a) Surface map of Retina_A (b) Texture map of Retina_A (c) OCT map of Retina_A

(d) Surface map of Retina_B (e) Texture map of Retina_B (f) OCT map of Retina_B

Fig. 3. Reconstructed 3-D shape of the retinal fundus (a)(d) Surface maps with an intersection line. (b)(e) Texture maps with intersection line in the same position. (c)(f) The enlarged 3-D OCT map of a blister area.

Fig. 3-(a)(d) shows different surfaces of 3-D retinal fundus that were obtained from several sequences of stereo images. Displayed depth values are scaled (by a factor of 30) to magnify the 3-D depth. Since the camera calibration is not performed, radial distortion and perspective effect still exist in the 3-D shape. The center of the estimated 3-D depth maps are very smooth with little noise and did not require a post processing step such as smoothing and regularization. The calculated depth values near the boundary are incorrect due to the occlusions.

In Fig. 3-(b)(e), the image texture is mapped over the 3-D depth map. The green lines, which respectively correspond to the line of the surface images on the left, display the 3-D depth of the texture. The 3-D shape of the blister and the optic disc are accurately estimated.

Fig. 3-(c)(f) presents OCT maps of the 'blisters'. 3-D OCT maps are extracted from data exported by the OCT machine using custom software (OCTOR, Doheny Image Reading Center). The size and degree of the 'blister' areas closely match the OCT data.

4.2 Combination of OCT and 3-D Shape Model

The reconstruction of 3-D surfaces is one of the most difficult processes in medical imaging because of the difficulty in accessing ground truth measurements. One possible source of ground truth data is obtained in this study from OCT. However, because OCT data and fundus images are acquired by different machines, it is hard to align the data to the 3-D shape data. Therefore, for evaluation, we manually matched several feature positions from the OCT to the 3-D depth map to align them to obtain a transformation matrix T. Differences between the 3-D point in the transformed OCT data, $T(X)$, and its nearest neighbor point in the 3-D depth map, Y, were calculated as an evaluation criterion.

$$E(X,Y,T) = \frac{1}{N_X} \sum_{\mathbf{x}_i \in X} \left\| T(\mathbf{x}_i) - \arg\min_{\mathbf{y}_j \in Y} \left\| T(\mathbf{x}_i) - \mathbf{y}_j \right\| \right\| \tag{4}$$

where N_X is the number of the points in OCT data.

We evaluated the performance of different stereo matching algorithms using the evaluation codes from the Middlebury database [15] for SSD, dynamic programming, scanline optimization, and graph cut algorithms. The best parameters for each image were selected for each method. We implemented normalized cross-correlation and mutual information. Table 1 demonstrates that, in our study, mutual information outperformed other methods in terms of accuracy. Mutual information generated little noise and is highly accurate, even in areas with low textures.

Table 1. Difference between OCT data and 3-D shape in pixel metric. For each method, parameters with the best performance for each image are selected.

	Retina_A	Retina_B	Retina_C
Sum of Squared Difference (SSD)	0.282563	0.510051	1.352426
Dynamic Programming (DP)	0.285503	1.142079	0.673890
Scanline Optimization (SO)	0.332272	0.626138	1.777096
Graph Cut (GC)	0.290186	0.515139	1.708525
Normalized Cross Correlation (NCC)	0.159760	0.437529	0.258479
Mutual Information (MI)	**0.152694**	**0.328937**	**0.245442**

5 Conclusion

Reconstructing the 3-D shape of the human retina from a stereo pair of fundus images is a challenging task because of intensity variation, the paucity of features in the macular area, and the plane-like surface of the retina. In this paper, we proposed a method for 3-D reconstruction of the retinal surface in an attempt to overcome such difficult cases. Specifically, our method employs a plane+parallax approach for robust epipolar geometry estimation and utilizes the mutual information criteria for

estimating a dense stereo map. The proposed method has been tested on numerous stereo sequences and validated by aligning OCT data with reconstructed 3-D depth data. Techniques such as this can be implemented easily in a telemedicine system which may have great utility in addressing the epidemics of eye disease that are emerging around the world due to diabetes, macular degeneration and glaucoma.

Acknowledgements

This research was partially supported by the National Institute of Health (NIH) under grant No. R21 EY015914-01, the Doheny Eye Institute, and the James H. Zumberge Faculty Research and Innovation Fund.

References

1. T.E. Choe, I. Cohen, “Registration of Multimodal Fluorescein Images Sequence of the Retina,” ICCV 2005, pp. 106-113, 2005
2. T.E. Choe, I. Cohen, G. Medioni, “3-D Shape Reconstruction of Retinal Fundus,” CVPR 2006, 2006
3. J. Gluckman and S. K. Nayar. Rectifying transformations that minimize resampling effects. In IEEE Proceedings of Computer Vision and Pattern Recognition, Kauai, December 2001.
4. C. Harris and M. Stephens, "A combined corner and edge detector", Proc. Alvey Vision Conf., Univ. Manchester, pp. 147-151, 1988.
5. H. Hirschmüller, “Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information,” CVPR 2005, Volume 2, 20-26 June 2005 Page(s):807 – 814
6. J. Kim, V. Kolmogorov, R. Zabih, “Visual correspondence using energy minimization and mutual information”, Ninth IEEE International Conference on Computer Vision, 2003. Proceedings, 13-16 Page(s):1033 - 1040 vol.2, Oct. 2003
7. V. Kolmogorov and R. Zabih, “Computing visual correspondence with occlusions using graph cuts,” in International Conference on Computer Vision, 2001.
8. R. Kumar, P. Anandan and K. Hanna, "Shape Recovery From Multiple Views: A Parallax Based Approach," DARPA IU Workshop, Monterey, Calif., Nov. 1994
9. H.C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. Nature, 293:133–135, Sept 1981.
10. D. Lowe, “Distinctive image features from scale invariant keypoints,” IJCV, Vol. 60, No. 2, pp. 91-110, 2004
11. A.G. Podoleanu, J.A. Rogers, D.A. Jackson, “Three dimensional OCT images from retina and skin”, Vol. 7, No. 9, Optic Express, Oct 2000
12. D. Scharstein and R. Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” IJCV, vol. 47, no. 1-3, pp. 7–42, April 2002.
13. J. Sun, N.N Zheng, H. Shum, “Stereo Matching Using Belief Propagation,” IEEE Trans. on PAMI, Vol. 25, No. 7, July 2003
14. P.A. Viola and W.M. Wells, III. Alignment by maximization of mutual information. International Journal of Computer Vision, 24(2):137–154, September 1997.
15. Webpage of Middlebury stereo database and source code: *www.middlebury.edu/stereo*

Comparing the Similarity of Statistical Shape Models Using the Bhattacharya Metric

K.O. Babalola[1], T.F. Cootes[1], B. Patenaude[2], A. Rao[3], and M. Jenkinson[2]

[1] Division of Imaging Science and Biomedical Engineering, University of Manchester, M13 9PT, UK
{kola.babalola/tim.cootes}@manchester.ac.uk
[2] FMRIB Centre, John Radcliffe Hospital, University of Oxford, OX3 9DU, UK
{mark/brian}@fmrib.ox.ac.uk
[3] Visual Information Processing, Department of Computing, Imperial College London, SW7 2BZ, UK
ar17@doc.ic.ac.uk

Abstract. A variety of different methods of finding correspondences across sets of images to build statistical shape models have been proposed, each of which is likely to result in a different model. When dealing with large datasets (particularly in 3D), it is difficult to evaluate the quality of the resulting models. However, if the different methods are successfully modelling the true underlying shape variation, the resulting models should be similar. If two different techniques lead to similar models, it suggests that they are indeed approximating the true shape change. In this paper we explore a method of comparing statistical shape models by evaluating the Bhattacharya overlap between their implied shape distributions. We apply the technique to investigate the similarity of three models of the same 3D dataset constructed using different methods.

1 Introduction

Statistical Shape Models (SSMs) are parametric representations of the shape and variation in the shape of a class of objects, learnt using a suitably representative training set. They are suited to medical structures as they allow the variation of the shape of the structure(s) being modelled to be captured. This proves useful for model fitting to unseen structures (segmentation), and investigations into the changes in shapes of anatomical structures due to disease (morphometric analysis).

The construction of statistical shape models requires that a correspondence is established across the training set. To construct SSMs of 3D structures, establishing correspondence manually is time consuming and subject to errors. Therefore, there is a burgeoning field of research into automatic methods for obtaining correspondence e.g. [1],[2]. In the case where the shapes of the structures being modelled are represented as points, the statistical shape model is a point distribution model (PDM) [3]. This is a linear model of the n-dimensional space in which each member of the training set is represented as a vector. Different methods of establishing correspondence for the same training set, represented by the same number of points, will result in different distributions in the n-dimensional space. Therefore, given the same training set represented by

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 142–150, 2006.

the same number of points, the different methods of establishing correspondence give rise to different SSMs.

Each such SSM captures information about the true variation in shape of the underlying dataset, but will also be influenced by the method used to obtain correspondences. It is important to be able to understand the true underlying variation in the data, especially for medical applications (where one may be looking for changes due to disease). It is often impractical to manually annotate enough data to construct models, so methods are required to confirm that the models resulting from automatic methods are reliable. One approach is to use two or more different model building methods - if they all result in similar models, one can begin to have confidence that the models are a true representation of the underlying variation. However, this requires a method of comparing the similarity of models which is the subject of this paper. The fact that two models are similar does not necessarily mean they are right, and methods of measuring model quality are still needed.

SSMs have been compared indirectly by evaluating their compactness and specificity [1],[4], and the quality of registration used in model construction has been evaluated using overlap measures [5]. In this paper we propose a method of quantitatively evaluating the similarity between two models. SSMs give compact representations of the probability distribution functions (PDFs) from which training examples are assumed to be drawn. Thus each SSM implies a PDF, and two models can be compared by evaluating the overlap between their implied PDFs. Such a measure gives insight into the different ways of constructing SSMs.

The Bhattacharya metric [6] has been shown to be suitable for comparing the similarity of two probability distributions. Thacker et. al. [7] have explored its application to model similarity in 1D and for parameter estimation. In the following, we describe the Bhattacharya measure and how we apply it to comparing the similarity of SSMs. Its application to 3D SSMs in their standard form is not possible because of memory requirements. An important contribution of this paper is the application of singular value decomposition to reduce the dimensionalities involved. Details of this are given in the Appendix at the end of the paper.

2 Bhattacharya Measure and Applicability

2.1 Bhattacharya Measure for Overlap of Two *n*-variate Gaussian Distributions

The Bhattacharya measure for two distributions is given by

$$B = \int \sqrt{p_1(\mathbf{x})p_2(\mathbf{x})} \quad d\mathbf{x} \tag{1}$$

The measure gives the overlap for two distributions p_1 and p_2 as a value between 0 (no overlap) and 1 (exact overlap) and is symmetric i.e. $B(p_1, p_2) = B(p_2, p_1)$. If $p_1(\mathbf{x})$ and $p_2(\mathbf{x})$ are two n-variate normal distributions with means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, and covariance matrices $\mathbf{V}_1$ and $\mathbf{V}_2$, then

$$p_i(\mathbf{x}) = |\mathsf{det}(\mathbf{V}_i)|^{-\frac{1}{2}}(2\pi)^{-\frac{N}{2}} \times \mathsf{exp}\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^{\mathsf{T}}\mathbf{V}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right) \tag{2}$$

The Bhattacharya overlap B(p_1,p_2) is then

$$B = 2^{\frac{N}{2}} |\mathsf{det}(\mathbf{V}_1)|^{-\frac{1}{4}} |\mathsf{det}(\mathbf{V}_2)|^{-\frac{1}{4}} |\mathsf{det}(\mathbf{V}_1^{-1} + \mathbf{V}_2^{-1})|^{-\frac{1}{2}} \times \quad \mathsf{exp}\left(-\frac{1}{4}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^{\mathsf{T}}(\mathbf{V}_1 + \mathbf{V}_2)^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\right) \tag{3}$$

2.2 Application to PDMs

A PDM is a linear model of the form (see [3])

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{Pb} + \mathbf{r} \tag{4}$$

Where $\mathbf{x}$ is a *shape vector* formed by concatenation of the coordinates of the points used to represent an instance of the class of shapes being modelled and $\boldsymbol{\mu}$ is the mean shape vector. $\mathbf{P}$ is a matrix whose columns are the eigenvectors of the covariance matrix, $\mathbf{b}$ is a vector of weights allowing instantiation of a particular set of values for $\mathbf{x}$, and $\mathbf{r}$ is a vector of residuals not explained by the model.

If we assume the original data was drawn from a Gaussian distribution, then each b_i is distributed as a zero mean Gaussian with variance λ_i (the eigenvalue of the original covariance matrix of the training data).

Given a model constructed by applying PCA to the shape vectors of a training set, we can estimate the covariance matrix from the model as

$$\mathbf{V} = \mathbf{P}^T \Delta \mathbf{P} + \sigma_r^2 \mathbf{I} \tag{5}$$

Δ is a diagonal matrix giving the variance of each parameter b_i and $\mathbf{I}$ is a square identity matrix. σ_r^2 represents the *variance of residuals*. It is given by:

$$\sigma_r^2 = \sigma_{\gamma_{min}}^2 + \left(\frac{\sum_{i=t+1}^{t_{max}} \lambda_i}{n-t}\right) \tag{6}$$

Each dimension in each of the distributions being compared has an element of noise associated with it. This is expressed by the first term in equation 6 and its value is determined empirically (see section 4.1). In applying PDMs to image search and morphometric analysis, not all the modes of the model are necessarily used as the first few modes usually contain most of the variation. In the case where not all modes are used, the mean of the variance of the modes left out (second term of equation 6) is included in determining the residual variance. In such cases t represents a *cutoff* in the number of modes of variation of the model, and t_{max} is the maximum number of modes of variation in the model. The value of σ_r^2 should be less than the lowest value of b_i of the modes used in the comparison.

2.3 Implementation Details

Calculating the Bhattacharya overlap involves computing the inverse and determinants of "large" matrices, which require huge amounts of computer memory and can lead to

numerical instability unless handled with care. The size of the covariance matrices are determined by the number of points used to represent the shapes. For n points in 3D the covariance matrix is of size $3n \times 3n$.

To allow application of the Bhattacharya overlap to 3D SSMs we use *singular value decomposition* (SVD) to move the $3n \times 3n$ covariance matrix into a smaller $m \times m$ matrix, where $m \leq 2\times$ number of training examples (see appendix for details). The Bhattacharya overlap is obtained using the smaller $m \times m$ matrix. In the present case we reduced our covariance matrices from $(76, 776 \times 76, 776)$ to (74×74). Results from 2D have shown that the values obtained for the Bhattacharya overlap with and without the SVD stage are equivalent.

3 Building the Statistical Shape Models

A training set consisting of magnetic resonance images of 37 subjects was used to build the 3D statistical shape models. The dataset contained T1 structural MR scans of the head of each subject at 1mm x 1.5mm x 1mm resolution. For each subject there was a corresponding image in which a number of cortical and subcortical structures had been labelled.

Three methods were used to build 3D SSMs of 10 subcortical structures (lateral ventricles, caudate, putamen, accumbens, pallidum, thalamus, amygdala, hippocampi, brain stem and the fourth ventricle). The first method was based on B-Spline free form deformations (FFD) [2], the second on composition of grid based diffeomorphic warps [8] and the last used a deformable surface mesh [9].

A reference subject was chosen by identifying the member of the training set that was most similar to all other subjects within the set. An initial mesh for each of the labelled structures in the reference subject was obtained using the deformable mesh. This point-set was used to build the SSM for each method. Therefore, there were three SSMs of the same training set built by different methods with the same number of points. The SSM produced by the mesh based method was used to investigate properties of the Bhattacharya metric, then the three models were compared with each other.

4 Experiments and Results

4.1 Calibrating the Metric

The Bhattacharya overlap returns a value between 0 and 1. In order to better interpret the values obtained when comparing models, we performed two different experiments to investigate how the metric degrades as model training data is corrupted. The model constructed using the mesh based method was used in these experiments.

In the first experiment random Gaussian noise was added in increasing amounts to vertices of the points of each member of the training set. The random noise was obtained from a Gaussian distribution with standard deviation expressed in fractions of the mean distance between connected vertices. The model constructed from the members of the training set with noise was compared to that constructed from the original training set.

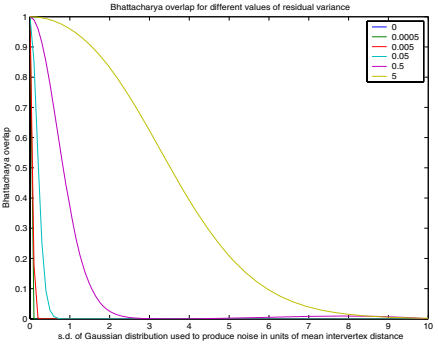


Fig. 1. Value of the Bhattacharya overlap as the amount of noise added to a 3D SSM is increased. The amount of noise is expressed in terms of the s.d. of the Gaussian distribution from which the noise values were sampled. The s.d. was in units of the mean distance between the connected vertices of the mesh. The values of $\sigma^2_{\gamma_{min}}$ for each plot are shown in the legend.

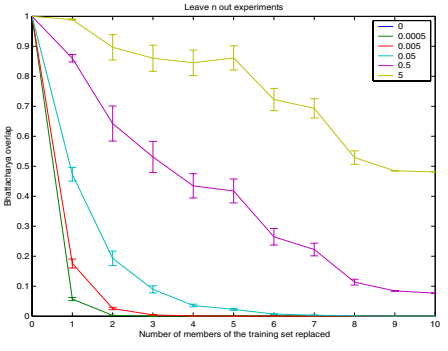


Fig. 2. Value of Bhattacharya overlap during leave *n* out experiments. The values for $\sigma^2_{\gamma_{min}}$ are shown in the legend

In the second experiment the members of the training set were incrementally replaced (*leave-n-out*). The training set was divided into two parts. 27 members were used to construct a SSM. The members of the 27 member training set were incrementally replaced by randomly picking members from the 10 member training set. The SSM constructed from this new set was compared with that of the original 27 member training set. For each data point 10 random choices from the left out members were used.

Figure 1 shows that the Bhattacharya overlap decreases as the amount of noise added to the model increases. The resulting curve has a sigmoid shape. Figure 2 shows that the Bhattacharya overlap decreases as the number of the 27 member training set replaced increases.

The results from the random noise experiment allows a relationship to be defined between the Bhattacharya overlap and the similarity of two models. For a particular value of the measure, the amount of noise required to achieve a similar difference from the initial distribution can be estimated.

Table 1. Values of the Bhattacharya overlap for comparing the first three modes of three SSMs of the same dataset. $\sigma^2_{\gamma_{min}}$ was set at 5.

Models Compared	Bhattacharya Overlap
FFD vs Compositional Warp	0.166
FFD vs Mesh	0.085
Mesh vs Compositional Warp	0.337

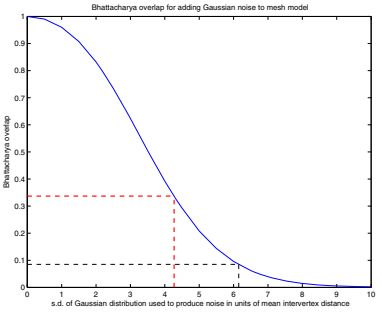


Fig. 3. Variation of Bhattacharya overlap with noise for Mesh model. The intercept marked by the red (thick) line shows the overlap with the Compositional warp model is equivalent to adding noise from a Gaussian distribution with s.d. 4.25 times the mean intervertex distance. That for the overlap with the FFD model (6.15 times the intervertex distance) is marked by the thin line.$\sigma^2_{\gamma_{min}}$ was set at 5.

4.2 Quantitative Model Similarity Results

Table 1 lists the values obtained for the Bhattacharya overlap when comparing the different SSMs. These values were obtained using $\sigma^2_{\gamma_{min}} = 5$. To give some physical meaning to the values, they are marked on a plot of the Bhattacharya overlap with noise for the Mesh model for $\sigma^2_{\gamma_{min}} = 5$ (figure 3).

4.3 Qualitative Analysis of Models

The mean shapes of the three models are shown in Figure 4. The figure also shows colour-mapped images of the difference between the mean shapes of the FFD and the compositional warp models relative to the mesh model. This shows that the mean shape of the mesh model is more similar to that of the compositional warp model than it is to the mean shape of the FFD model.

In the PDM equation 4, the eigenvectors contained in $\mathbf{P}$ each define a *mode of variation*, with eigenvectors corresponding to the largest eigenvalues describing the most significant modes of variation. To visualise the differences captured by the modes, instances of the shape can be generated between limits determined by each b_i. We instantiated shapes between $+2.5\sqrt{\lambda_2}$ and $-2.5\sqrt{\lambda_2}$ of the first three modes of variation for each of the different models. However, we could not draw definite conclusions on which models were more similar by observing these modes of variation.

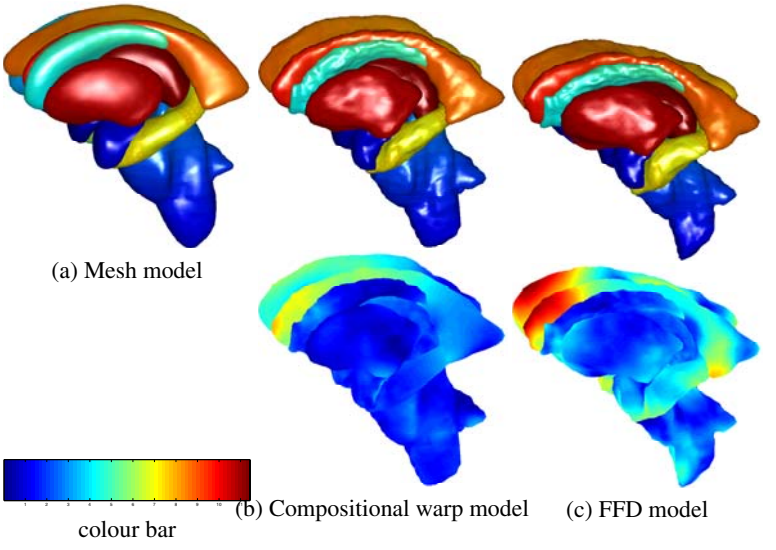


Fig. 4. The mean shapes of the three models. In subfigures (b) and (c) the difference in Euclidean distance between the corresponding points in the Mesh model was used to colour-map the surfaces in the 2nd row. The relationship between the colours and the Euclidean distance (in mm) is shown in the colour bar

5 Discussion

The results have shown that the Bhattacharya overlap can be used as a measure of the similarity of two SSMs. We have quantified the value of the metric in terms of the amount of Gaussian noise added to a model. The residual variance acts as a regularisation term allowing the sensitivity of the measure to be adjusted.

Results from the permutation experiments suggest that the metric could also be used to evaluate the generalisation ability of a model. This is the ability of the model to adequately represent the training set. We would expect a model that is representative of a class of shapes to be relatively insensitive to the actual members of the training set. However, as the results of figure 2 show this is not the case for the models considered here. This is a common problem in medical applications - basically the numbers in the training set are too small.

We are pursuing several aspects of this work further. Evidently, $\sigma^2_{\gamma_{min}}$ plays an important role in determining the value obtained for the Bhattacharya overlap. We intend to investigate their relationship further and explore alternative approaches to calibrating the measure. We would also like to investigate decoupling the differences in the mean shapes of the two models from the differences in the modes of variation. Lastly, we intend to further explore the application of the Bhattacharya overlap to measuring the generalisation ability of 3D SSMs.

To conclude, we have presented a novel method of comparing statistical shape models. The method gives us a new approach for evaluating the quality of models built automatically from large datasets, particularly in the medical field.

Acknowledgements

This work was funded by the EPSRC under the IBIM project. David Kennedy of the Center for Morphometric Analysis, Boston, provided the MR images used.

References

1. Twining, C.J., et. al.: A unified information-theoretic approach to groupwise NRR and model building. In: LNCS (IPMI 2005 Proceedings). Volume 3565. (2005) 1–14
2. Rueckert, D., Frangi, A.F., Schnabel, J.A.: Automatic construction of 3-D statistical deformation models of the brain using NRR. IEEE TMI **22** (2003) 1014–1025
3. Cootes, T.F., et. al.: ASMs - their training and application. CVIU **61** (1995) 38–59
4. Styner, M., et. al.: Evaluation of 3d correspondence methods for model building. In: LNCS (IPMI 2003 Proceedings). Volume 2732. (2003) 63–75
5. Crum, W.R., Camara, O., Rueckert, D., Bhatia, K.K., Jenkinson, M., Hill, D.L.G.: Generalised overlap measures for assessment of pairwise and groupwise image registration and segmentation. In: (MICCAI proceedings). Volume 3749 of LCNS. (2005) 99–106
6. Bhattacharya, A.: On a measure of divergence between two statistical populations defined by their probability distributions. Bulletin of Calcutta Maths Society **35** (1943) 99–110
7. Thacker, N.A., Aherne, F., Rockett, P.I.: The bhattacharya metric as an absolute similarity measure for frequency coded data. Kybernetica **34** (1997) 363–368
8. Cootes, T.F., Twining, C.J., Taylor, C.J.: Diffeomorphic statistical shape models. In: Proceedings of the 15th British Machine Vision Conference, Norwich. (2004) 447–456
9. Patenaude, B., Smith, S., Jenkinson, M.: A bayesian cost function applied to model-based registration of sub-cortical brain structures. In: To appear in LCNS volume on Proceedings of the Workshop on Biomedical Image Registration (WBIR) 2006. (2006)

Appendix - Using SVD to Decrease Matrix Size

Given two models of training sets consisting of s training examples each with n landmarks defined in three dimensions :

$$\mathbf{x}_1 = \boldsymbol{\mu}_1 + \mathbf{P}_1\mathbf{b}_1 + \mathbf{r} \qquad and \qquad \mathbf{x}_2 = \boldsymbol{\mu}_2 + \mathbf{P}_2\mathbf{b}_2 + \mathbf{r} \tag{7}$$

$\mathbf{x}$ and $\mathbf{r}$ are column vectors of length $3n$, $\mathbf{P}$ is of size $(3n \times s-1)$, and $\mathbf{b}$ is a row vector of length $(s-1)$.

We create a $(3n \times 2s)$union space of vectors and perform an SVD:

$$\mathbf{U}\mathbf{W}\mathbf{V}^T = SVD(\boldsymbol{\mu}_1|\boldsymbol{\mu}_2|\mathbf{P}_1|\mathbf{P}_2) \tag{8}$$

$\mathbf{U}$ is a $(3n \times 2t)$ matrix whose columns form basis vectors in the union space. The components of the models in the union space can be obtained by projecting using $\mathbf{U}$:

$$\mathbf{U}^T\mathbf{x}_i = \mathbf{U}^T\mu_i + \mathbf{U}^T\mathbf{P}_i\mathbf{b}_i + \mathbf{U}^T\mathbf{r} \qquad \text{and} \qquad p_i(\mathbf{x}) = p_i^{(U)}(\mathbf{U}^T\mathbf{x}) \times p_i^{(\bar{U})}(\mathbf{C}^T\mathbf{x}) \tag{9}$$

where $\mathbf{C}$ is the complement of $\mathbf{U}$. Then

$$\begin{aligned} B(p_1, p_2) &= \int \sqrt{p_1^{(U)} p_1^{(\bar{U})} \quad p_2^{(U)} p_2^{(\bar{U})}} \quad d\mathbf{x} \\ &= \int \sqrt{p_1^{(U)} p_2^{(U)}} \cdot \sqrt{p_1^{(\bar{U})} p_2^{(\bar{U})}} \quad d\mathbf{x} \\ &= B(p_1^{(U)}, p_2^{(U)}) B(p_1^{(\bar{U})}, p_2^{(\bar{U})}) \end{aligned} \tag{10}$$

If we assume that the variance of the noise in both models is the same, then the distributions in the null space of $\mathbf{U}$, $p_1^{(\bar{U})}() = p_2^{(\bar{U})}()$ and $B(p_1^{(\bar{U})}, p_2^{(\bar{U})}) = 1$. Therefore,

$$B(p_1, p_2) = B(p_1^{(U)}(\mathbf{U}^T \mathbf{x_1}), p_2^{(U)}(\mathbf{U}^T \mathbf{x_2})) \tag{11}$$

and the $(2s \times s - 1)$ matrix $\mathbf{U}^T \mathbf{P}$ can be used to estimate a $(2s \times 2s)$ covariance matrix, instead of using the $(3n \times s - 1)$ matrix $\mathbf{P}$ to estimate a $(3n \times 3n)$ covariance matrix.

Improving Segmentation of the Left Ventricle Using a Two-Component Statistical Model

Sebastian Zambal, Jiří Hladůvka, and Katja Bühler

VRVis Research Center for Virtual Reality and Visualization,
Donau-City-Strasse 1, 1220 Vienna, Austria
http://www.vrvis.at

Abstract. Quality of segmentations obtained by 3D Active Appearance Models (AAMs) crucially depends on underlying training data. MRI heart data, however, often come noisy, incomplete, with respiratory-induced motion, and do not fulfill necessary requirements for building an AAM. Moreover, AAMs are known to fail when attempting to model local variations. Inspired by the recent work on split models [1] we propose an alternative to the methods based on pure 3D AAM segmentation. We interconnect a set of 2D AAMs by a 3D shape model. We show that our approach is able to cope with imperfect data and improves segmentations by 11% on average compared to 3D AAMs.

1 Introduction

Correct segmentation of the left ventricle (LV) in cardiac MRI short-axis images is very important for further diagnosis of the heart's function. Manual segmentation is very time-consuming and therefore efforts are taken to automate this process to the largest possible extent.

Although variants of Active Appearance Models (AAMs) [2] have been applied to this problem [3,4,5], a robust and completely automatic segmentation still poses a challenge.

Several problems appear in the context of shape- and appearance-based modeling of the left ventricle:

- The appearance of papillary muscles and trabeculae varies irregularly from patient to patient. Especially in slices close to the heart's apex these fuzzy anatomical structures are hard to model statistically. The necessary requirement for building an AAM – that the underlying training data has a gaussian distribution – is not satisfied in the apical/apex region.
- Often, only a part of the left ventricle is captured. Training data consisting of poorly corresponding volumes decreases the quality of the resulting model.
- Respiratory-induced motion introduces shifts of neighboring slices and so leads to irregular local variances in the data.
- Principal Component Analysis (PCA) involved in the build leads to a model where variances apply to the complete model. This makes it hard to handle local variations, e.g. different brightness of individual slices.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 151–158, 2006.

In this paper we propose a new and robust approach to segmentation of the left ventricle. By linking a set of 2D AAMs with a simple 3D shape model we perform model matching on a global and local context iteratively. With this approach we handle the above list of problems.

This paper is structured as follows. In section 2 we review related work on statistical models of shape and appearance in context of LV segmentation. In section 3 we introduce a two-component model of the left ventricle. In section 4 we outline how the two-component model is matched to unseen data. Results are presented in section 5 and a conclusion is drawn in section 6.

2 Statistical Modeling of the Left Ventricle

Our approach benefits from the combination of both Active Shape Models (ASMs) [6] and Active Appearance Models (AAMs) [2]. ASMs are statistical models of shape that can be used for identification and segmentation of known objects in unknown images. AAMs are a direct extension of ASMs and incorporate not only information about shape but also model the underlying texture. Variations of both methods have successfully been applied to segmentation in medical imaging [4,7,8].

Statistical models of shape and appearance are created from a finite number of training data sets. In other words, the model is only able to represent and thus segment new data sets that are similar to the ones in the training set. Increasing the size of the training set is a rather limited way to improve the generalization ability of the model. By adding local deformability the generalization ability can be increased explicitly. The statistical model is used to find a rough initial segmentation which is refined by restricted local deformations [8].

Several authors have discussed local and global deformability of a model. Independent Component Analysis (ICA) [9,10] leads to modes of variation which have a rather local effect compared to the global modes of variation derived from the conventionally used PCA. To deal with local variations of a model it can explicitly be split into sub-models. The split is achieved either with the help of human expertise or by means of optimization based on the Minimal Description Length (MDL) [1].

A great problem that arises in segmentation of short-axis MRI images of the left ventricle is respiratory-induced motion. Such artifacts appear as slight displacements of spatially neighboring slices. Handling of these displacements is crucial to achieve both a correct model and correct segmentations. A recently proposed approach [11,12] first detects the left ventricle in one or multiple slices. In a second step image alignment is used to compensate the shifts of neighboring slices. Finally a 3D AAM search is performed to achieve the final segmentation.

3 A Two-Component Model of the Left Ventricle

Inspired by the idea of modeling global and local features separately [1,8] we propose to combine a set of local 2D AAMs with a global 3D shape model. The purpose of the 2D AAMs is to precisely match the well articulate slices of the heart's

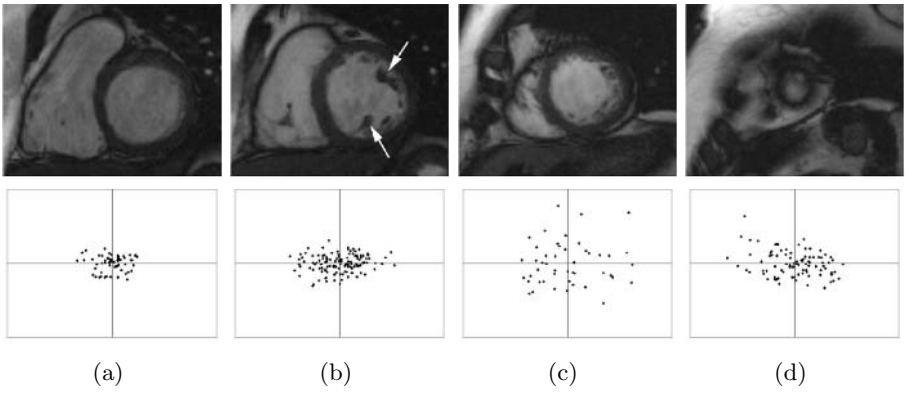


Fig. 1. Typical examples of (a) basal, (b) mid (the arrows indicate papillary muscles), (c) apical, and (d) apex slices (top) and the first two principal components of texture after PCA (bottom)

base. The purpose of the 3D shape model is twofold: to propagate the position and size of the basal slices to apical ones and to keep the global shape characteristics plausible. In the following we describe the two components of our model.

3.1 Component 1: A Set of 2D Active Appearance Models

In a first step we consider the given training data, i.e. texture and annotation, slice-wise. From base to apex we identify the following four classes of slices:

Basal slices (fig. 1(a)) reside close to the base of the heart and do not contain papillary muscles.

Mid slices (fig. 1(b)) clearly show papillary muscles (arrows).

Apical slices (fig. 1(c)) contain trabeculae and papillary muscles, which are represented by irregular texture.

Apex slices (fig. 1(d)) show the very apex of the heart and contain highly irregular texture.

We performed PCA on the shape-normalized texture vectors of all slices in the training set. Scatter-plots of the first two principal components are shown in figure 1 for the according classes of slices. We observe that basal and mid slices show a relatively compact gaussian distribution and are thus well suited for linear statistical modeling using PCA. This is not the case for apical and apex slices, whose texture is much more irregular.

With a manual classification of slices we build a set of four individual 2D AAMs. This proceeding has the advantage that even incomplete MRI data sets can be included in the training set and used for building the first component of the model.

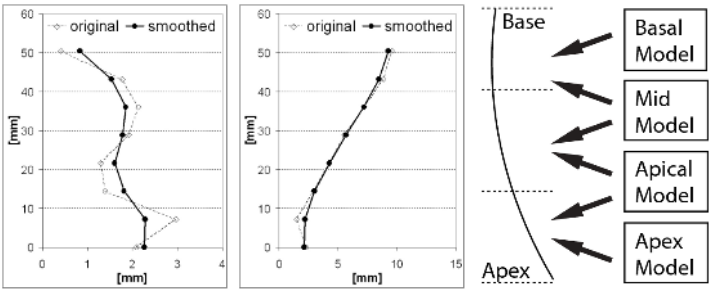


Fig. 2. The original and smoothed centerlines projected to XZ (a) and YZ (b) plane. (c) The assignment of sub-models to sections of the global shape model.

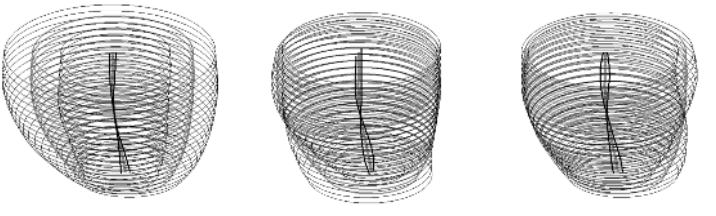


Fig. 3. The first (a), second (b), and third (c) modes of variation of the global shape model. The plots show the mean shape together with the largest possible deformations.

3.2 Component 2: A Global Shape Model

The intention of the global model is to represent the ventricle's shape in an overall simplified way. Its task is to assure a valid relative placement and scaling of the four 2D AAMs.

The core of the global model is a centerline connecting the 2D AAMs in 3D. For the landmark points of each slice of the annotated training set a centroid is calculated. The assembly of these centroids forms the centerline. Figures 2(a) and 2(b) show such a centerline for one of the annotated data sets. It can be seen that respiratory-induced shifts of neighboring slices appear in the data. We use a gaussian filter to smooth the centerlines and to reduce respiratory motion artifacts in the training set.

Additionally to the centroids we incorporate the radii of slices into the global model. For each slice a radius as the mean distance of shape points to the centroid is calculated. This results in three variables that approximately describe the ventricle's shape on one slice: x/y-coordinates and radius.

To build a statistical shape model, an equal number of corresponding landmarks has to be placed in every training example. We interpolate centroid positions and radii to get a total of 20 evenly-spaced slices. With the three features per slice this results in a total of 60 features included in the global shape model.

Before PCA can be applied, the individual training examples have to be aligned. Since we model not only the positions of centroids but also the radii we can not directly apply 3D Procrustes analysis. We rather align the center lines with respect to x/y-coordinates. The radii are thus not affected by the alignment. Figure 3 illustrates the first three modes of variation of the global shape model.

3.3 Combining Local Models and Global Model

In order to benefit from both components the local models have to be linked to the global model. As some data sets are delivered without apex or basal slices, we compensate this missing information by attaching two 2D AAMs to each slice. For example, the top most slice is assigned the basal model and the mid model. The better matching model is kept while the other one is ignored. Figure 2(c) shows the assignment of 2D AAM sub-models to the global shape model.

4 Matching the Two-Component Model

In the previous section we have outlined the idea of splitting a statistical model of the left ventricle into four parts and described how those parts are coupled over a global model. In this section we explain how the model iteratively is matched to unseen data.

4.1 Matching the Local Sub-models

We switch between two strategies. First a local matching of 2D AAMs is carried out for individual slices by standard AAM search [2]. In theory it should be possible to match all 2D AAMs in this way to get a valid global segmentation. While a valid match is obtained in basal and mid areas the 2D AAM search very often fails in the apical and apex slices. This is where the global model comes into play.

4.2 Updating the Global Model

Using the global component we propagate information from the better fitting slices to badly matching slices: The root mean square (RMS) texture error is calculated for all slices and the best 80% are used to update the global shape model. This update is done analogously to ASM search [6].

4.3 Iterative Global and Local Matching

After the global model has been updated the local AAMs are aligned to it. The search proceeds iteratively switching between 2D AAM search and update of the global model. In this way divergence due to bad initialization in critical slices is avoided and the overall segmentation gets improved. The loop of 2D and 3D updates is repeated until convergence.

The resulting segmentation still suffers from the lack of local variability. Further improvements are achieved performing additional 2D AAM search steps with position updates restricted by a constant maximum displacement.

Finally respiratory-induced shifts are compensated. The slices are shifted such that the centroids of the 2D AAMs align to the global model's centroid coordinates. Compared to the work of Stegmann [11] our motion compensation thus benefits from prior knowledge that is encoded in the global shape model.

5 Results

We evaluated the described method with a set of 32 different short-axis studies. The quality of segmentation achieved with the two-component model was compared to that of 3D AAMs. Leave-one-out tests for all of the 32 data sets were performed. The average point-to-surface error (PSE) with respect to expert annotation was used to validate the segmentation quality. Figure 4 illustrates the measurements. Although the two-component model led to slightly worse results in 9 cases out of 32, the overall PSE got improved from an average of 2.20mm (3D AAM) to 1.96mm (the two-component model). Even though the average improvement by 0.24mm does not sound impressive, we emphasize that compared to the standard 3D AAM the two-component model performs better by 11%.

In figure 5 we provide a visual comparison for the largest improvement in terms of the average PSE which could be achieved for data set 10. It shows the result obtained with a standard 3D AAM (fig. 5(a)) and the improved segmentation achieved with the two-component model (fig. 5(b)).

Figure 6 demonstrates the motion compensation performed with the two-component model. A section of data set number 10 is depicted in figure 6(a). Figure 6(b) shows the data set after matching and aligning the slices to the centerline of the global shape model.

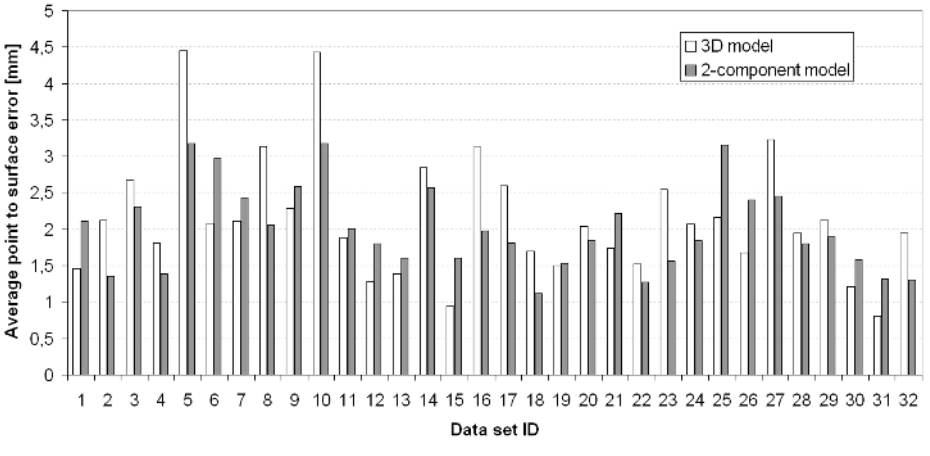
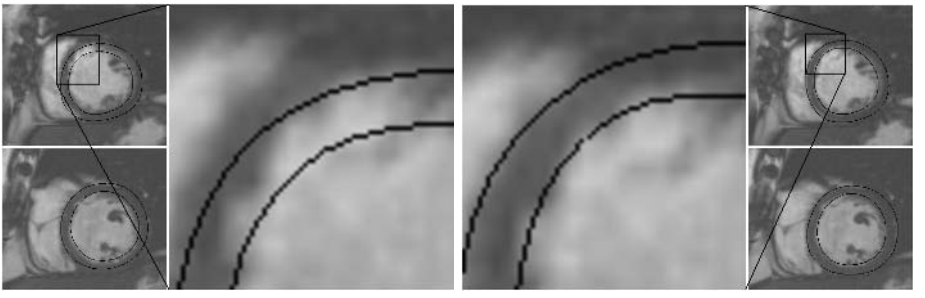
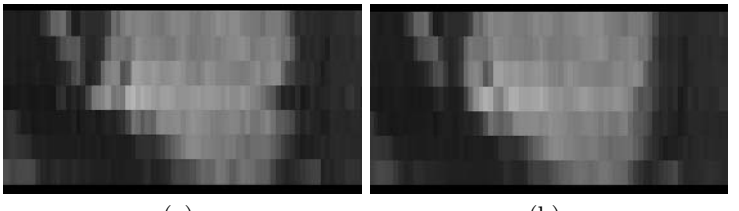


Fig. 4. Point-to-surface error measured in leave-one-out tests for all training data sets



(a) (b)

Fig. 5. Results of matching data set 10 with 3D AAM (a) and two-component model (b)



(a) (b)

Fig. 6. Motion compensation of data set 10 carried out by the two-component model. The original data (a) and the automatically corrected data (b) are shown.

Between 10 and 20 iterations are necessary for the two-component model matching algorithm to converge. The whole segmentation process typically takes a few seconds on standard PC hardware.

6 Conclusion, Discussion, Future Work

We have introduced a new approach to 3D segmentation of the left ventricle from short-axis MRI images by interconnecting a set of 2D AAMs with a simple 3D shape model. The two-component model is more robust than a 3D AAM since local irregularities such as respiratory-induced motion and different intensities of gray values in individual slices can be handled.

Incomplete training data sets that do not contain the entire LV from base to apex pose a severe problem for 3D AAMs. If such data sets are present in the training set the correspondence between them is ill-posed. The outcoming model allows for shrinking within its modes of variation. The topmost or the bottommost slices thus often remain unsegmented. This disadvantage is avoided with the two-component model which allows the use of incomplete data sets as training examples. Since slices are assigned to individual 2D AAM sub-models even a single annotated 2D short-axis slice can be added to the training set. As the two-component model attempts to match two neighboring 2D AAMs to each slice, segmentation of data sets that do not cover the entire LV is still possible.

We are convinced that the idea of combining different types of models over a global shape model is promising for other medical segmentation problems. Especially if local anatomical abnormalities (e.g., a tumor) appear, models consisting of multiple components will likely lead to more precise results.

References

1. Langs, G., Peloschek, P., Donner, R., Bischof, H.: A clique of active appearance models by minimum description length. In: British Machine Vision Conference (BMVC). Volume 2. (2005) 859–868
2. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. In: European Conference on Computer Vision (ECCV). Volume 2. (1998) 484–498
3. Stegmann, M.B., Pedersen, D.: Bi-temporal 3D active appearance models with applications to unsupervised ejection fraction estimation. In: International Symposium on Medical Imaging. Volume 5747. (2005)
4. Mitchell, S.C., Lelieveldt, B.P.F., van der Geest, R.J., Bosch, J.G., Reiber, J.H.C., Sonka, M.: Multistage hybrid active appearance model matching: Segmentation of left and right ventricles in cardiac MR images. IEEE Transactions on Medical Imaging **20**(5) (2001) 415–423
5. Mitchell, S.C., Bosch, J.G., Lelieveldt, P.F., van der Geest, R.J., Reiber, J.H.C., Sonka, M.: 3D active appearance models: Segmentation of cardiac MR and ultrasound images. IEEE Transactions on Medical Imaging **21**(9) (2002) 1167– 1178
6. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models – Their training and application. Computer Vision and Image Understanding **61**(1) (1995) 38–59
7. Beichel, R., Bischof, H., Leberl, F., Sonka, M.: Robust active appearance models and their application to medical image analysis. IEEE Transactions on Medical Imaging **24**(9) (2005)
8. Taylor, C.J., Cootes, T.F.: Combining elastic and statistical models of appearance variation. In: European Conference on Computer Vision (ECCV). (2000) 149–163
9. Üzümcü, M., Frangi, A., Reiber, J., Lelieveldt, B.: The use of independent component analysis in statistical shape models. In: SPIE Medical Imaging. Volume 5032. (2003) 375–383
10. Suinesiaputra, A., Frangi, A.F., Üzümcü, M., Reiber, J.H.C., Lelieveldt, B.P.F.: Extraction of myocardial contractility patterns from short-axes MR images using independent component analysis. In: ECCV Workshops CVAMIA and MMBIA. (2004) 75–86
11. Stegmann, M.B., Ólafsdóttir, H., Larsson, H.B.W.: Unsupervised motion-compensation of multi-slice cardiac perfusion MRI. Medical Image Analysis **9**(4) (2005) 394–410
12. Stegmann, M.B., Larsson, H.B.W.: Motion-compensation of cardiac perfusion MRI using a statistical texture ensemble. In: Functional Imaging and Modeling of the Heart, FIMH 2003. Volume 2674. (2003) 151–161

An Approach for the Automatic Cephalometric Landmark Detection Using Mathematical Morphology and Active Appearance Models*

Sylvia Rueda and Mariano Alcañiz

Medical Image Computing Laboratory (MedICLab), Universidad Politécnica de Valencia, UPV/ETSIA, Camino de Vera s/n, 46022 Valencia, Spain
silruelo@degi.upv.es, malcaniz@degi.upv.es

Abstract. Cephalometric analysis of lateral radiographs of the head is an important diagnosis tool in orthodontics. Based on manually locating specific landmarks, it is a tedious, time-consuming and error prone task. In this paper, we propose an automated system based on the use of Active Appearance Models (AAMs). Special attention has been paid to clinical validation of our method since previous work in this field used few images, was tested in the training set and/or did not take into account the variability of the images. In this research, a top-hat transformation was used to correct the intensity inhomogeneity of the radiographs generating a consistent training set that overcomes the above described drawbacks. The AAM was trained using 96 hand-annotated images and tested with a leave-one-out scheme obtaining an average accuracy of 2.48mm. Results show that AAM combined with mathematical morphology is the suitable method for clinical cephalometric applications.

1 Introduction

Cephalometry means measurement of the head. Accordingly, a cephalometric analysis consists of characterizing distances and angles between significant structures in a x-ray image of the head. The whole process is based on the localization of cephalometric landmarks. These points correspond to specific locations on the radiograph and are accurately defined for hard, external soft and pharyngeal tissues [1]. Figure 1 shows the location of commonly chosen cephalometric points for several methods. Once located, lines are traced between the points to measure angles and distances, which are the cephalometric indices. After a comparison of these measurements with the norms stated for the same age, ethnicity and sex, a diagnosis can be established and a treatment forecast.

Manual positioning of landmarks is a tedious, time-consuming and error prone task. Depending on the quality of the radiograph and the experience of the user, the location will be slightly different. According to Dean et al. [2] and Geelem et al. [3], inter-expert variability during the positioning task varies from 3.3mm to 5mm and intra-expert variability is of 1mm.

* Radiographs provided by Y. Perrin, A. Sanpietro and the Eastman Dental Institute.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 159–166, 2006.

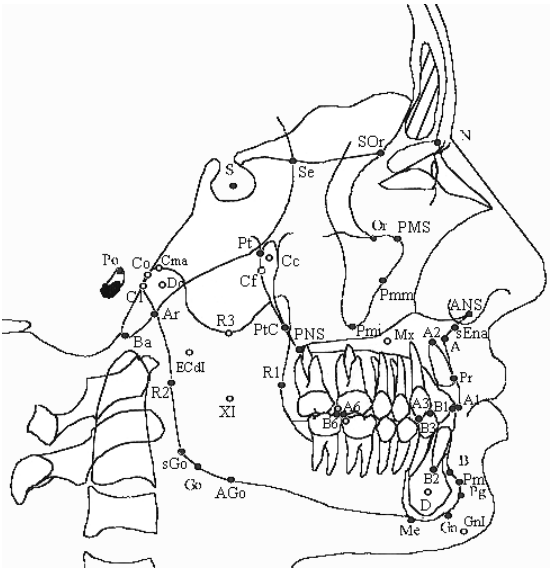


Fig. 1. Lateral cephalometric points of bony tissue for different methods of analysis

The orthodontist spends on average 15-20 minutes per analysis depending on the quality of the cephalogram, his/her experience and the number of points considered in the method selected. A fully automated system would detect landmarks, thus reducing the time required to carry out an analysis, improving the accuracy of landmark localization and decreasing errors due to expert subjectivity.

2 Previous Work

Several approaches have been proposed for the automatic detection of cephalometric landmarks. They can be classified into 3 groups:

Knowledge-based Systems. The first attempt was undertaken by Levy et al. [4], who applied enhancement filters and a knowledge-based line tracker to extract edges. Landmarks were located according to their geometric definitions. The method was tested on two high-quality x-ray images. 23 of 36 landmarks were located on good quality images. Parthasaraty et al. [5] improved the previous work by including a four level resolution pyramid to reduce time-processing. The testing was done in five images. Nine of the landmarks considered in that study, 58% were within ±2mm, 18% within ±1mm and 100% within ±5mm. These studies used few images for training and tested the method on the training set. In the presence of artifacts and bad quality images, these methods will not work and, therefore, cannot be used for clinical cephalometric systems.

Template Matching. Template matching methods implement a grey-level model around each landmark on each image in the training set. Then, they search

for a correspondence between the model and new images. Cardillo et al.[6] applied a pattern-matching algorithm based on mathematical morphology to detect landmarks. They considered 40 images in the training set. The testing was done on 20 images. Of 20 landmarks, 66% were within ±2mm. Grau et al. [7] improved on the work of Cardillo et al. by using a line detection module to search for the most significant lines. They used 20 images for training and 20 for testing and reported that 90% of the 17 landmarks tested were repeatedly located within ±2mm. Chen et al. [8] developed a method of neural networks and genetic algorithms to search for sub-images containing each of the cephalometric points. These methods consider the texture around each point. However, this appearance can vary from one image to another making local detection worse. Moreover, not enough images were used to represent the variability existing in cephalograms. In our experience, these methods, when used on images slightly different from the ones of the training set, are not sufficiently accurate for clinical applications.

Statistical Models. Statistical methods are one of the most suitable options for considering the huge amount of variability in cephalograms, as they take into account the variation of characteristics in the images. Hutton et al. [9] applied Active Shape Models (ASMs) to cephalometric point detection. They used 63 randomly selected cephalograms and tested the training set with a drop-one-out method. 13% of 16 landmarks were within ±1mm, 35% within ±2mm and 74% within ±5mm. The authors concluded that ASM did not give sufficient accuracy for landmark detection in clinical applications and could only be used as a good starting point for global landmark identification. ASM mainly considers the variation in shape present on the images. The intensity variations are only modeled in the profiles normal to the contour. Accordingly, to improve accuracy it is necessary to add more information to the system. The appearance of the radiographs contains useful information that should be taken into account.

3 Method

Active Appearance Models (AAMs), recently proposed by Cootes et al. [10][11], can model both shape and texture variability existing in a training set. To create a model we need a set of labeled images representative of the real variability of the object we want to segment. AAMs have proved to be powerful tools for medical image segmentation and understanding [12]. The more homogeneous the images of the training set are, the more robust the model will be. The problem of the images we work with in this research is that they can be greatly different and present huge luminosity inhomogeneity. Mathematical morphology can solve this; in particular, we will apply a top-hat transformation [13] on the images to extract light objects from an uneven background.

3.1 Selection of the Training Set

The images forming the training set must be representative of the variability existing in clinical cases. The data and the source should reflect the target

population for the algorithm that is being developed. In particular, it is necessary to consider: the anatomical and morphological variability of the human head, the variability of texture depending on image quality, the variability of structures present in a cephalogram (not all the radiographs have the same size and include the same structures), the variability of capturing the x-ray (i.e. double structures appearing because of the non-orthogonal position of the patient during capture of a lateral head film) and the variability of the source.

3.2 Luminosity Inhomogeneity Correction

Depending on the source, the cephalogram can present varied textures and thus be completely different from one another. Background cannot be separated from the rest due to its non-uniformity. Therefore, a purely preprocessing approach for intensity inhomogeneity correction is required to improve results when using AAM. Mathematical morphology [14], which is based on set theory, provides powerful tools for image analysis. Fundamental operations are erosion, dilation, opening and closing. An opening consists of an erosion followed by a dilation. A structuring element defines the size and the shape of the transformation to be done. In our case we will use a circular structuring element with a radius of 75 pixels. We use this size to extract the structures of interest correctly. The white top-hat $\rho(f)$ extracts bright structures and is defined as the difference between the image f and its opening $\gamma(f)$:

$$\rho(f) = f - \gamma(f) \tag{1}$$

Subtracting the open image from the original provides an image where removed objects stand out clearly. This technique is used to extract contrasted components with respect to the background. The result of applying this transformation can be seen in Fig. 2 (right column).

3.3 Training the Model

To generate a statistical model of shape and texture variation it is necessary to describe the shape and the texture of each training example. Therefore, we represent n landmark points, (x_i, y_i), for each image as a $2n$ vector, $\mathbf{x}$, where $\mathbf{x} = (x_1, \ldots, x_n, y_1, \ldots, y_n)^T$ will describe the shape of an object. The annotated training set is aligned into a common co-ordinate frame using a Procrustes Analysis [15]. Hence, we obtain the Point Distribution Model (PDM) for all the images of the training set. The mean shape is extracted and the appearance variation collected by establishing a piece-wise affine warp (based on the Delaunay triangulation) between each image of the training set and the mean shape. Next, the intensity is sampled from the shape-normalized images over the region covered by the mean shape. The resulting samples are then normalized to minimize the effect of global lighting variation and the texture (grey-levels) vector $\mathbf{g}$ is obtained. Finally, for both shape and texture, we perform a Principal Component Analysis (PCA) on the aligned training set to describe the shape

and appearance variations of the object. Often, shape and texture are correlated. Therefore, we can deduce a combined model of shape and texture with parameters $\mathbf{c}$, controlling the shape and texture at the same time. We get:

$$\mathbf{x} = \overline{\mathbf{x}} + \mathbf{Q}_s\mathbf{c} \ , \ \mathbf{g} = \overline{\mathbf{g}} + \mathbf{Q}_g\mathbf{c} \tag{2}$$

where $\overline{\mathbf{x}}$ and $\overline{\mathbf{g}}$ are the mean shape vector and the mean normalized grey-level vector, $\mathbf{Q}_s$ and $\mathbf{Q}_g$ are matrices describing the modes of variation derived from the training set and $\mathbf{c}$ the combined PCA parameters. New images of the object can be synthetically constructed from $\mathbf{c}$. More details about AAMs can be found in [10][11].

3.4 Segmenting New Images

We first place an initial template model over the unseen image. We then use a principal component multivariate linear regression model to generate new images to fit the unseen image in the best way. Once the process converges, a match can be declared. Finally, an optimization scheme accomplishes further refinement of the match.

3.5 Model Evaluation

The final step of the whole process consists of testing the performance of the algorithm. Care should be taken to test the algorithm in data that has not been used in the training.

In this case, we will use a common methodology called leave-one-out. In this approach, a set of M images (ground truth) is split in M different ways into a training set of size M-1 and a test set of size 1. For each of the M splits, training is done on the M-1 images and then testing is done on the remaining 1 item. Performance is then estimated as the average of the M tests.

4 Experimental Results

We applied the method described to automatically detect cephalometric landmarks. The open C++ source code AAM-API [17] was partially used in this study. We implemented the leave-one-out algorithm in Matlab to evaluate the model.

96 images, annotated by an expert, form the training set used in this study. Image selection fulfils the requirements stated in Sect. 3.1. This number of images is sufficient to include all the existing variability. Each image has a size of 780×1000 pixels and 256 grey-levels. 43 cephalometric landmarks were considered to assess most of the cephalometric analysis. We used a combination of scanned and digital x-ray images, coming from different sources, to consider different image qualities in the training set. Anatomical variability is assessed with good proportion between ages, different norm groups (Caucasian, black and oriental), incisor relationship (I,II or III) and dentition (mixed or permanent).

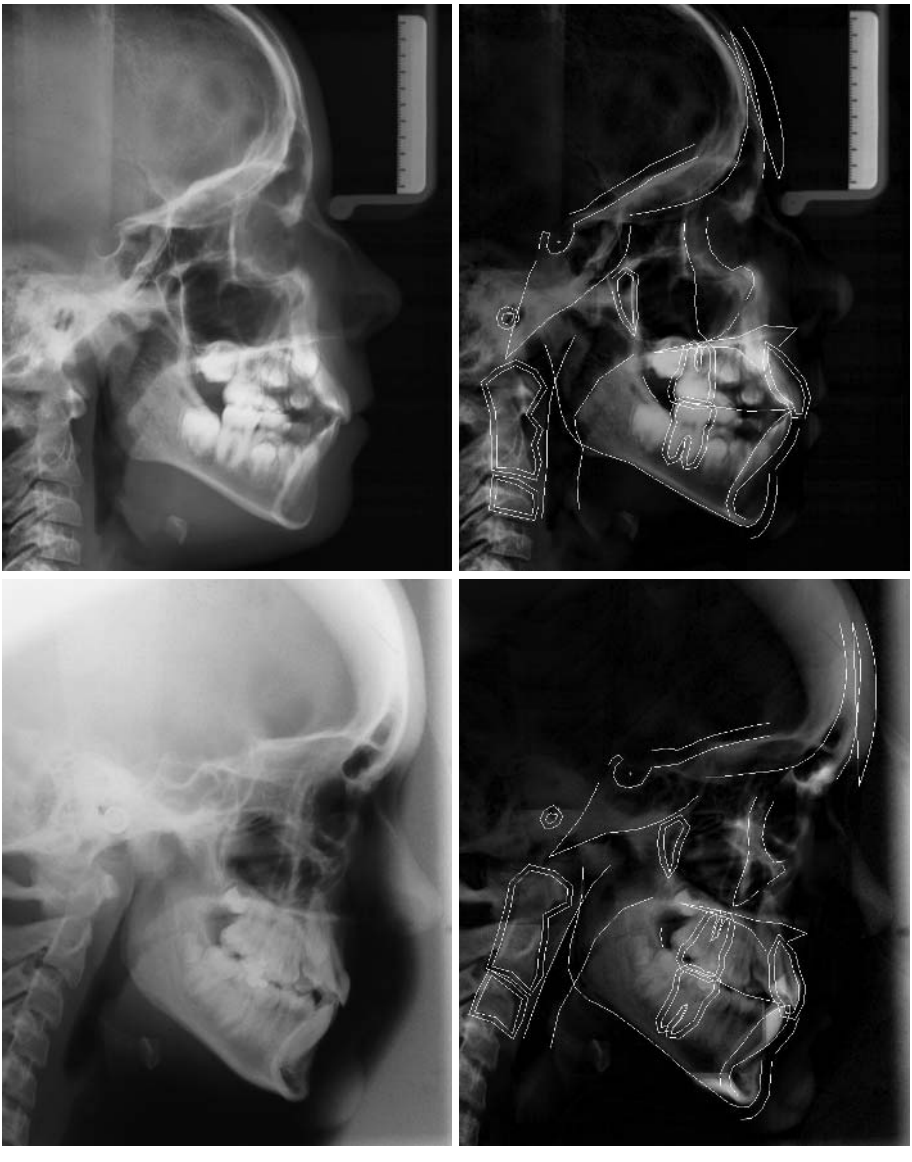


Fig. 2. Results of the search in a digital image (first row) and a scanned x-ray (second row) after luminosity inhomogeneity correction. We can appreciate that the original x-rays are very different and the intensity normalized images are more homogeneous.

We have created a template with 289 points. Points were divided into cephalometric, mathematical and pseudo-landmarks [15]. Double contours were added to some structures to improve the search giving more specificity to the model [16] (Fig. 2). The evaluation of the model was assessed using a leave-one-out method. We considered the initialization of the system to have failed if more than 60% of

the points have an error larger than 5mm. In these conditions, for 96 images used in this study, the automatic initialization failed in 9.37% of the cases. Figure 2 shows the results of different quality cephalograms. The first row assesses the result in a digital x-ray whereas the second row shows the result in a scanned x-ray with high luminosity variation, black regions and rotation of the head. For both, the segmentation works well and with high precision. Table 1 assesses the accuracy of the system for the main landmarks considered. The names of the landmarks correspond to locations in Fig. 1.

Table 1. Results and comparisons of the main landmarks

	mean (mm)	std (mm)	≤ 2mm (%)	≤ 3mm (%)	≤ 5mm (%)		mean (mm)	std (mm)	≤ 2mm (%)	≤ 3mm (%)	≤ 5mm (%)
Na	2.31	1.78	56.32	75.86	91.95	**SOr**	2.16	1.33	49.42	80.46	96.55
Se	2.04	1.12	52.87	80.46	100	**S**	2.29	1.26	39.08	70.11	97.70
Or	2.05	1.23	57.47	73.56	97.70	**Po**	3.66	2.12	18.39	42.53	78.16
PMS	2.24	1.25	48.28	75.86	96.55	**Co**	2.47	1.35	37.93	64.37	94.25
Ba	2.70	1.79	37.93	67.82	91.95	**Ar**	2.27	1.41	44.83	74.71	96.55
Pt	2.19	1.22	47.13	78.16	98.85	**PtC**	2.14	1.37	55.17	78.16	97.70
Cf	2.02	1.18	54.02	85.06	98.85	**R1**	1.95	1.02	60.92	85.06	100
B	2.19	1.26	43.68	80.46	97.70	**Pm**	1.97	1	56.32	85.06	98.85
Pg	1.83	1.06	57.47	89.66	98.85	**Gn**	1.58	1.12	73.56	86.21	98.85
Me	1.59	1.07	70.11	86.21	100	**Go**	3.88	2.41	26.44	43.68	66.67
R2	2.46	1.43	41.38	71.26	91.95	**LIT**	1.52	0.90	77.01	94.25	100
LIR	1.55	0.92	67.82	94.25	100	**PNS**	2.67	1.42	36.78	68.97	91.95
ANS	2.12	1.28	55.17	78.16	96.55	**A**	2	1.41	67.82	83.91	94.25
UIT	1.82	1.85	75.86	87.36	94.25	**UIR**	1.98	1.43	65.52	82.76	95.40

On average, for all the cephalometric landmarks considered, the system has a precision of 2.48mm and an average standard deviation of 1.66mm. Considering that inter-expert variability varies from 3.3mm to 5mm, these results are to date the best option for clinical applications. In this study, 50.04% of the landmarks were located within 2mm, 72.62% within 3mm and 91.44% within 5mm.

5 Conclusions

In this paper, we have shown that AAM achieves precise, automatic detection of cephalometric landmarks that takes into account the variability existing in shape and texture. This leads to the clinical applicability of automatic landmark identification. Previous work in this field used few images, tested their methods in the training set or did not select a consistent training set that well characterized the variability of this type of images. None of the previous studies undertaken had achieved clinical cephalometric analysis. Our method has proved to be useful to locate a considerable number of the landmarks with a high precision rate for clinical cephalometric applications. Image homogeneization and the use of double contours for some structures improve AAM results when applied to cephalometric radiographs.

References

1. Rakosi, T.: An Atlas and Manual of Cephalometric Radiology, London Wolfe Medical Publications (1982)
2. Dean, D., Palomo, M., Subramanya, K., et al.: Accuracy and precision of 3D cephalometric landmarks from biorthogonal plain-film x rays. SPIE Med. Imag. 3335, (1998) 50-58
3. Geelem, W., Wenzel, A., Gotfredsen, E., Kruger, M. and Hansson, L.G.: Reproducibility of cephalometric landmarks in conventional film, and hardcopy and monitor-displayed images obtained by the storage phosphor technique. Eur. J. Orthod. 20, (1998) 331-340
4. Lévy-Mandel, A., Venetsanopoulos, A. and Tsotsos, J.: Knowledge-based landmarking of cephalograms. Comput. Biomed. Res. 19, (1986) 282-309
5. Parthasaraty, S., Nugent, S., Gregson, P. et al.: Automatic landmarking of cephalograms. Comput. Biomed. Res. 22, (1989) 248-269
6. Cardillo, J. and Sid-Ahmed, M.A.: An image processing system for locating craniofacial landmarks. IEEE Trans. Med. Imag. 13, (1994) 275-289
7. Grau, V., Alcañíz, M., Juan, M.C., Monserrat, C. and Knoll, C.: Automatic Localization of Cephalometric Landmarks. J. biomed. inform. 34, (2001) 146-156
8. Chen, Y.T., Cheng, K.S. and Liu, J.K.: Improving Cephalogram Analysis through Feature Subimages Extraction. IEEE Eng. in Med. Biol. Mag. 18, (1999) 25-31
9. Hutton, T.J., Cunningham, S. and Hammond, P.: An Evaluation of Active Shape Models for the Automatic Identification of Cephalometric Landmarks. Eur. J. Orthod. 22, (2000) 499-508
10. Cootes, T.F., Edwards, G.J. and Taylor, C.J.: Active Appearance Models. Proc. European Conference on Computer Vision, vol.**2** (1998) 484-498
11. Cootes, T.F. and Taylor, C.J.: Statistical Models of Appearance for Computer Vision, Report (2004)
12. Cootes, T.F. and Taylor, C.J.: Statistical models of appearance for medical image analysis and computer vision, Proc. SPIE Med. Imag. (2001)
13. Sonka, M. and Fitzpatrick, J.M.: Handbook of Medical Imaging, vol.2.Medical Image Processing and Analysis, SPIE Press (2000)
14. Sonka, M., Hlavac, V. and Boyle, R.: Image Processing, Analysis, and Machine Vision, Second Edition (1999) Brooks/Cole Publishing Company
15. Dryden, I.L. and Mardia, K.V.: Statistical Shape Analysis, Wiley (1998)
16. Stegmann, M.B.: Active Appearance Models - Theory, Extensions & Cases, Master Thesis IMM-EKS-2000-25 (2000)
17. Stegmann, M.B., Ersboll, B.K. and Larsen, R.: FAME - a flexible appearance modelling environment. IEEE Trans. Med. Imag. (2003)

Automatic Segmentation of Jaw Tissues in CT Using Active Appearance Models and Semi-automatic Landmarking*

Sylvia Rueda[1], José Antonio Gil[1], Raphaël Pichery[2], and Mariano Alcañiz[1]

[1] Medical Image Computing Laboratory (MedICLab), Universidad Politécnica de Valencia, UPV/ETSIA, Camino de Vera s/n, 46022 Valencia, Spain
silruelo@degi.upv.es, jgil@dsic.upv.es, malcaniz@degi.upv.es
[2] Ecole Supérieure d'Ingénieurs de Luminy, Marseille, France
rpichery@esil.univ-mrs.fr

Abstract. Preoperative planning systems are commonly used for oral implant surgery. One of the objectives is to determine if the quantity and quality of bone is sufficient to sustain an implant while avoiding critical anatomic structures. We aim to automate the segmentation of jaw tissues on CT images: cortical bone, trabecular core and especially the mandibular canal containing the dental nerve. This nerve must be avoided during implant surgery to prevent lip numbness. Previous work in this field used thresholds or filters and needed manual initialization. An automated system based on the use of Active Appearance Models (AAMs) is proposed. Our contribution is a completely automated segmentation of tissues and a semi-automatic landmarking process necessary to create the AAM model. The AAM is trained using 215 images and tested with a leave-4-out scheme. Results obtained show an initialization error of 3.25% and a mean error of 1.63mm for the cortical bone, 2.90mm for the trabecular core, 4.76mm for the mandibular canal and 3.40mm for the dental nerve.

1 Introduction

Dental implants are titanium roots placed in the bone of the jaw to support replacement teeth. Preoperative planning systems are commonly used for oral implant surgery. The dentist or surgeon placing the implant needs to determine if the quality and quantity of bone is sufficient for long-term function and stability.

The body of the mandible is formed by a hard exterior, the cortical bone, and a soft spongy inside, the trabecular core. Teeth are fixed to the jaw by their roots and nerved to the mandibular canal. The canal contains the mandibular nerve which has to be avoided to prevent permanent or temporary lip numbness. This nerve runs from the area behind the wisdom teeth, passes under the molars and emerges under the skin of the face in the region where the premolar teeth are or used to be (Fig. 1).

* This research has been partially supported by 3DENT Sistemas Dentales, Spain.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 167–174, 2006.

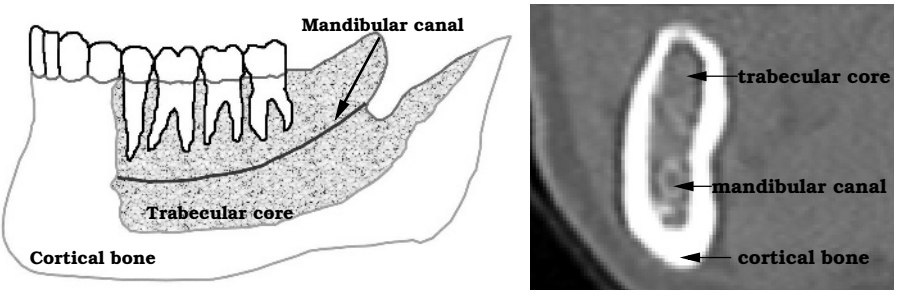


Fig. 1. Mandible structure: diagram and CT sectional view

Oral implant surgery uses different views on CT images to plan the intervention. Our work aims to segment cortical bone, trabecular core and the mandibular canal in sectional views (Fig. 1, right) because of their simplicity and the coherence existing between nearby slices. Depending on the patient, the shape of the bone can be very different and present mandible bone reabsorption. Moreover, the boundary of the trabecular core is not always well defined. We attempt to detect this region roughly because the mandibular canal is located inside. The canal is not always distinguishable in the trabecular zone because of the similarity of intensities in this region. Furthermore, these images present artifacts and sometimes teeth which make segmentation difficult.

2 Previous Work

We aim to automate the segmentation of cortical bone, trabecular core and the mandibular canal in CT images. Most of the studies carried out in the area of implantology has focused on 3D mandible reconstruction from projections [1] [2] avoiding the segmentation of tissues. Few studies have been found about mandible tissues segmentation. In most cases, the method was based on threshold and needed the intervention of an expert. For example, Fütterling et al. [3] developed an approach to automate the segmentation process of jaw tissues. They concluded that it was not possible to reliably detect any border between the cortical bone and the trabecular core without massive interaction. Stein et al. [4] proposed an approach to the detection of the mandibular canal using Dijkstra's algorithm to find the path and a deformable model to build the tubular structure. The process required the initialization of an expert.

De Bruijne et al. [5] obtained good results in the segmentation of tubular structures using Active Shape Models (ASMs). ASMs adapt the shape of the object to be segmented according to the statistical information of a training set previously annotated by an expert.

Previous work has shown that the problem to solve is complex and often requires expert interaction. Threshold techniques have proved to be inefficient to separate cortical bone and trabecular core. Nonetheless, the use of ASM to segment tubular structures gave good results. Therefore, if we want more accuracy

we have to consider texture information. Active Appearance Models is an extension of ASMs that takes into account texture and shape variability. We will use this method to segment mandibular tissues automatically.

3 Method

Active Appearance Models (AAMs), recently proposed by Cootes et al. [6][7], can model both shape and texture variability existing in a training set. To create a model we need a set of labeled images representative of the real variability of the object we want to segment. AAMs have proved to be powerful tools for medical image segmentation and understanding [8]. The whole process followed is summarized in Fig. 2.

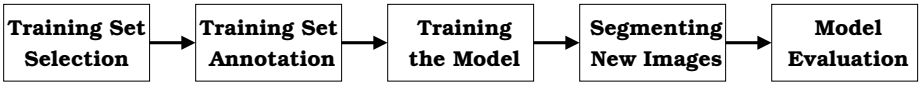


Fig. 2. Methodology

3.1 Selection of the Training Set

The images forming the training set must be representative of the variability existing in real cases. The data and the source should reflect the target population for the algorithm that is being developed.

3.2 Semi-automatic Annotation of the Training Set

Because of the high variability of shape and the few anatomic landmarks existing in these images, manual landmarking, with high correspondence of points between images, will be difficult, tedious and error prone. The best solution is to automatize the process as much as possible. We use threshold techniques to find the contour of each structure. First, we find the external contour of the cortical bone (Fig. 3.a). Second, we define five mathematical landmarks (points of high curvature) on the cortical contour (Fig. 3.b). Third, we find the contour of the trabecular core (Fig. 3.c). Then, we locate the dental nerve, which is the only anatomical landmark, in the center of the mandibular canal (Fig. 3.d). Finally, we select the radius of the canal (Fig. 3.e).

Once this is done, we automatically define pseudo-landmarks equally spaced between the anatomical and mathematical landmarks previously located. To describe each structure we use a large number of landmarks. As stated in [9], duplicating the structure gives more specificity to the model and more accurate segmentation. The annotation for each structure can be seen in black in Fig. 4. Each black point is a landmark.

The cortical bone (Fig. 4.a) is described with 30 landmarks and a double contour of 28 points. The structure is open to avoid the presence of teeth. The

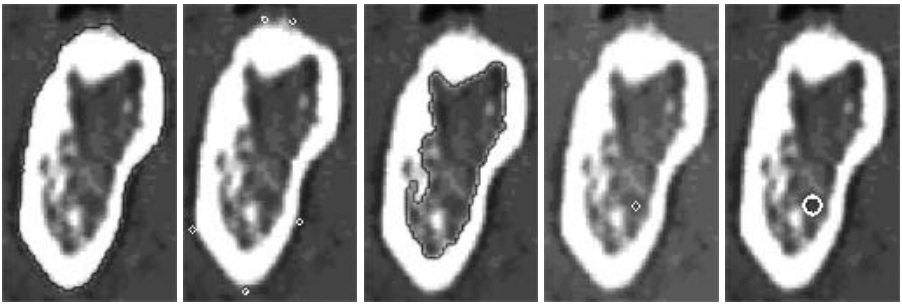


Fig. 3. Semi-automatic landmarking steps

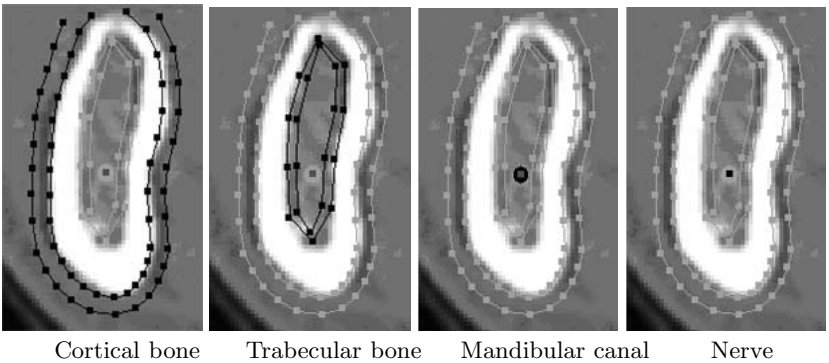


Fig. 4. Model landmarking

trabecular core is described with 10 landmarks and a double contour of 10 points. This structure has high variability and we use few points to avoid the adaptation of the shape to details. The mandibular canal uses 8 landmarks and the dental nerve only one.

3.3 Training the Model

To generate a statistical model of shape and texture variation it is necessary to describe the shape and the texture of each training example. Therefore, we represent n landmark points, (x_i, y_i), for each image as a $2n$ vector, $\mathbf{x}$, where $\mathbf{x} = (x_1, \ldots, x_n, y_1, \ldots, y_n)^T$ describes the shape of an object. The annotated training set is then aligned in a common co-ordinate frame using a Procrustes Analysis. Hence, we obtain the Point Distribution Model (PDM) for all the images of the training set. The mean shape is extracted and the appearance variation collected by establishing a piece-wise affine warp (based on the Delaunay triangulation) between each image of the training set and the mean shape. Next, the intensity is sampled from the shape-normalized images over the region covered by the mean shape. The resulting samples are then normalized to minimize the effect of global

lighting variation and the texture (grey-levels) vector $\mathbf{g}$ is obtained. Finally, for both shape and texture, we perform a Principal Component Analysis (PCA) on the aligned training set to describe the shape and appearance variations of the object. Often, shape and texture are correlated. Therefore, we can deduce a combined model of shape and texture with parameters $\mathbf{c}$, controlling the shape and texture at the same time. We obtain:

$$\mathbf{x} = \overline{\mathbf{x}} + \mathbf{Q}_s \mathbf{c} \tag{1}$$

$$\mathbf{g} = \overline{\mathbf{g}} + \mathbf{Q}_g \mathbf{c} \tag{2}$$

where $\overline{\mathbf{x}}$ and $\overline{\mathbf{g}}$ are the mean shape vector and the mean normalized grey-level vector, $\mathbf{Q}_s$ and $\mathbf{Q}_g$ are matrices describing the modes of variation derived from the training set and $\mathbf{c}$ the combined PCA parameters. New images of the object can be synthetically constructed from $\mathbf{c}$. More details about AAMs can be found in [6][7]. The whole process is summarized in Fig. 5.

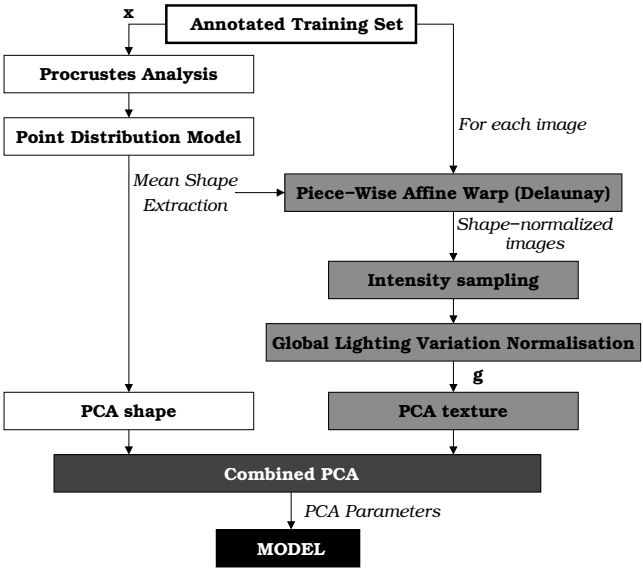


Fig. 5. Building an Active Appearance Model

3.4 Segmenting New Images

We first place an initial template model over the unseen image. We then use a principal component multivariate linear regression model to generate new images to fit the unseen image in the best way. Once the process converges, a match can be declared. Finally, an optimization scheme accomplishes further refinement of the match.

3.5 Model Evaluation

The final step of the whole process consists of testing the performance of the algorithm. Care should be taken to test the algorithm in data that has not been used in the training.

In this case, we will use a common methodology called leave-N-out. In this approach, a set of M images (ground truth) is split in N different ways into a training set of size M-N and a test set of size N. For each of the N splits, training is done on the M-N images and then testing is done on the remaining N. Performance is then estimated as the average of the N tests.

For each image segmented, we compare the result with the corresponding ground truth image. We calculate the distance between the structures segmented and those previously annotated, which gives the average error for each image and for each structure. There are two ways of measuring this distance: point-to-point (pt.pt) or point-to-curve (pt.crv). The pt.pt distance measures the Euclidean distance between each corresponding landmark, whereas the pt.crv distance measures the shortest distance to the curve in the neighborhood of the corresponding landmark. Pt.crv distance is more representative of the global segmentation because it does not evaluate each isolated point. We therefore based our evaluation on the pt.crv distance.

4 Results

The method described to segment cortical bone, trabecular core and the mandibular canal was applied to 215 CT images which were selected from 62 patients coming from different sources in order to cover the variability existing in real cases. The entire model was described by 87 landmarks annotated on each image of the training set. Semi-automatic landmarking was undertaken to give more accuracy to the system and assure that all the intermediate points were really equally spaced. The open C++ source code AAM-API[10] was partially used in this study. We created a tool with OpenCV libraries in C++ to annotate the images semi-automatically and to build AAM models automatically. We developed the leave-4-out to evaluate the models.

To calculate the optimized parameters in shape, we fixed the texture variance at 40% and the combined variance at 95%. We then took into account the results obtained and fixed the shape variance at 25% to calculate the optimized parameters in texture. The results can be seen in Fig. 6.

The best configuration is for 25% of the shape variance and 50% of the texture variance. For the optimal configuration, results in mean and standard deviation for each structure are shown in Table 1.

For a mean pt.crv distance lower than 5mm for at least 90% of the points for each image or structure, a good initialization can be defined. Accordingly, 90% of cortical bones, 66% of trabecular cores, 70% of mandibular canals and 72% of dental nerves were found correctly for all the 215 images. An example of good segmentation can be seen in Fig. 7.

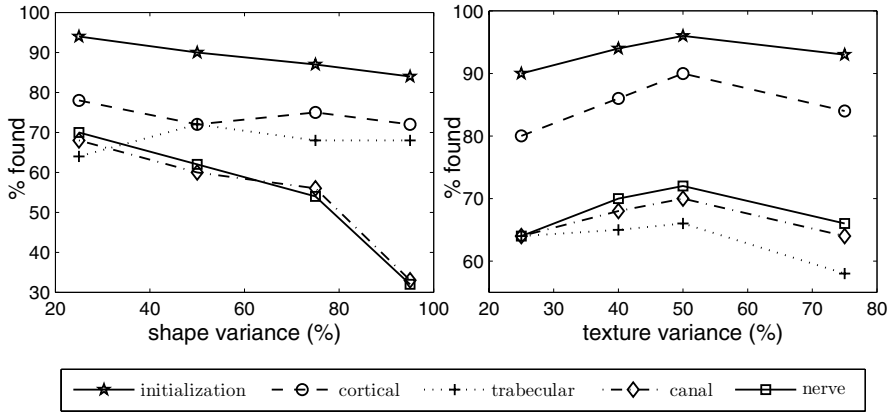


Fig. 6. Influence of shape and texture variance considered

Table 1. Results for the optimal configuration

	mean (mm)				std (mm)			
	Cortical	Trabecular	Canal	Nerve	Cortical	Trabecular	Canal	Nerve
pt.crv	1.63	2.90	4.76	3.40	1.63	2.16	0	1.11

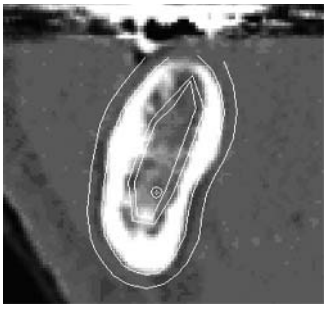


Fig. 7. Successful fitting

We do not expect more accuracy in the trabecular core because this structure has a high shape variability and we annotated the training set to avoid trabecular details during the segmentation.

5 Conclusion

AAMs perform accurate and automatic segmentation of cortical bone, trabecular core, mandibular canal and dental nerve taking into account all the variability of real cases. None of the previous work undertaken had achieved a real, automatic jaw tissues segmentation. In the research presented in this paper, semi-automatic

landmarking was developed to improve the AAM method and decrease annotation time.

References

1. Moreno, A., Bloch, I., Dutreuil, J. and Deval, M.: 3D Reconstruction from Radiologic Projections applied to Dental Implantology and to oral-facial surgery, ENST Report (2004).
2. Verstreken, K.: An Image-Guided Planning System for Endosseous Oral Implants. IEEE Transactions on Medical Imaging, vol.**17** (1998), 5.
3. Fütterling, F., Klein, R., Straber, W., Weber, H.: Automated Finite Element Modeling of a Human Mandible with Dental Implants. 6-th International Conference in Central Europe on Computer Graphics and Visualization, 1998.
4. Stein, W., Hassfeld, S., and Muhling, J.: Tracing of thin tubular structures in computer tomographic data. Computer Aided Surgery vol.**3** (1998), 83-88
5. DeBruijne, B.M., Van Ginneken, B., Niessen, W.J., Viergever, M.A., Wiro, J.N.: Adapting Active Shape Models for 3D Segmentation of Tubular Structures in Medical Images, Information Processing in Medical Imaging Vol.**2732** (2003). Lecture Notes in Computer Science, 136-147
6. Cootes, T.F., Edwards, G.J. and Taylor, C.J.: Active Appearance Models. Proc. European Conference on Computer Vision, vol.**2** (1998) 484-498
7. Cootes, T.F. and Taylor, C.J.: Statistical Models of Appearance for Computer Vision, Report (2004)
8. Cootes, T.F. and Taylor, C.J.: Statistical models of appearance for medical image analysis and computer vision, Proc. SPIE Medical Imaging (2001)
9. Stegmann, M.B.: Active Appearance Models - Theory, Extensions & Cases, Master Thesis IMM-EKS-2000-25 (2000)
10. Stegmann, M.B., Ersboll, B.K. and Larsen, R.: FAME - a flexible appearance modelling environment. IEEE Transactions on Medical Imaging (2003)

Morphometric Analysis for Pathological Abnormality Detection in the Skull Vaults of Adolescent Idiopathic Scoliosis Girls

Lin Shi[1,2], Pheng Ann Heng[1,2], Tien-Tsin Wong[1,2], Winnie C.W. Chu[3], Benson H.Y. Yeung[4], and Jack C.Y. Cheng[4]

[1] Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China
{lshi, pheng, ttwong}@cse.cuhk.edu.hk
[2] Shun Hing Institute of Advanced Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China
[3] Department of Diagnostic Radiology and Organ Imaging, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China
winnie@ruby.med.cuhk.edu.hk
[4] Department of Orthopaedics and Traumatology, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China
{byeung, jackcheng}@cuhk.edu.hk

Abstract. In this paper, we present a comprehensive framework to detect morphological changes in skull vaults of adolescent idiopathic scoliosis girls. To our knowledge, this is the first attempt to use a combination of medical knowledge, image analysis techniques, statistical learning tools, and scientific visualization methods to detect skull morphological changes. The shape analysis starts from a reliable 3-D segmentation of the skull using thresholding and math-morphological operations. The gradient vector flow is used to model the skull vault surface, which is followed by a spherically uniform sampling. The scale-normalized distances from the shape centroid to sample points are defined as the features. The most discriminative features are selected using recursive feature elimination for support vector machine. The results of this study specify the skull vault surface changes and shed light on building the evidence of bone formation abnormality in AIS girls.

1 Introduction

Adolescent idiopathic scoliosis (AIS) is a cause-unknown complicated 3-D spine deformity appearing most commonly in girls between 10 and 16 years old. The current treatment of AIS is still unsatisfactory because the etiopathogenesis remains unclear [1]. To enable evidence-based treatments and prognosis of curve progression, research on AIS has been carried out from various aspects, e.g. neuromotor mechanisms, hormonal system, and genetics [2]. In previous research studies carried out by our center, we have found persistently low bone mineral density [3,4,5,6,7,8] and abnormal skeletal growth in AIS during peri-pubertal

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 175–182, 2006.

development, which includes both peripheral [9] and axial skeleton [10,11]. Compared with age-matched controls, the longitudinal growth of the vertebral bodies in AIS is disproportionate and faster and mainly occurs by endochondral ossification. In contrast, the circumferential growth by membranous ossification is slower in both the vertebral bodies and pedicles [10]. We have hypothesized that there is generalized abnormal systemic growth in AIS and there might be associated abnormal growth and ossification of whole skull, which comprises the endochondral and membranous ossification process. In this work, we sought to analyze skull vault morphologies, which might bring out inspiring findings of ossification abnormality of the skull in AIS girls.

In a typical shape analysis application, the first step is the acquisition of the 2-D or 3-D shape from the data; it usually refers to image segmentation. Then shape descriptors are extracted to quantitatively characterize the shape. Finally, the shape difference accounts for the classification can be inferred from statistical analysis performed on these numerical descriptors.

Shape Extraction. There is limited work on segmenting skulls from MRI data, since the signals of bony structures is very weak in MRI. In recent work of Dogdas et al., [12], a human skull and scalp segmentation method is proposed based on thresholding and morphological operations.

Shape Description. Driven by various applications, shape description methods include parametric methods, e.g., spherical harmonic transform [13], and non-parametric methods, e.g., distance transform [14]. Snakes have been used to represent an object boundary as a parameter curve or surface. Xu et al. proposed the gradient vector flow (GVF) snake [15], which addresses the problems of the short capture range and inability to track at boundary concavity.

Statistical Analysis. The determination of discriminative descriptors that separates the two groups with least error can be interpreted as a feature selection problem in the machine learning domain. Feature selection methods can be classified into three categories [16], i.e., filter, wrapper, and embedded approaches. Various feature selection schemes have been proposed for the popular large margin classifier, support vector machine (SVM). Guyon et al. [16] introduced SVM-RFE, a backward elimination method for gene selection using SVM.

Figure 1 shows the framework of the proposed skull vault analysis scheme. The implementation details are introduced in the remainder of this paper.

2 Skull Vault Extraction and Description

2.1 Skull Segmentation

Reliable 3-D skull segmentation is an indispensable prerequisite for further feature extraction and statistical analysis. However, defining the skull in MRI is a challenging problem because of the weak magnetic resonance signals in the bony structures. We adopt a recently proposed effective strategy for skull segmentation in MRI data [12], which uses simple thresholding and math-morphological operations but achieves results close to CT data, especially in the range of skull vault. Before segmentation, the images are first preprocessed using the anisotropic

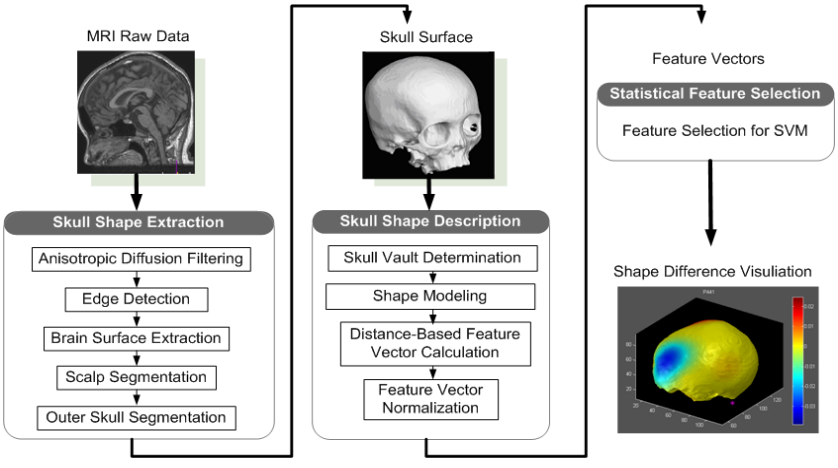


Fig. 1. The framework of skull vaults difference learning proposed in this paper

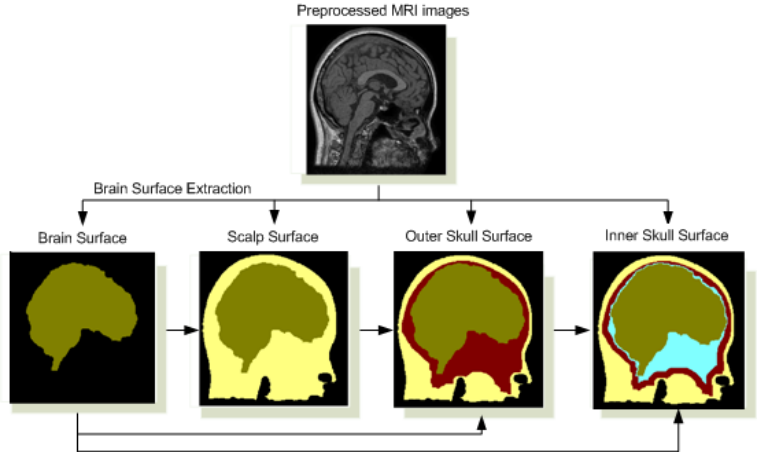


Fig. 2. The process of skull segmentation using thresholding and 3-D math-morphological operations

diffusion filter, which smoothes the volume while preserving sharp edges. The brain is extracted using the BSE algorithm [17]. The segmented brain is used as an inner bound for further skull segmentation. Thereafter, the scalp, the outer bound of the skull, is segmented using a combination of thresholding and math-morphological operations. Then the skull surface is found between outlines of brain surface and scalp using thresholding and morphological closing and opening. Figure 2 gives a summary of the skull segmentation process.

2.2 Skull Vault Description

In our particular application, we are interested in localizing the discriminative features in the skull vaults between AIS patients and normal controls, thus *local*

descriptors are preferred, which implies straightforward correspondences with the spatial distribution of the skull vault.

Skull Vault Determination. The skull vault is normally measured in anthropometry as the upper part of the skull determined by the plane passing through the three well-defined anatomical landmarks: bilateral frontozygomatic synchondrosis and the most superior posterior point of the external auditory meatus, as shown in Figure 4 (a). In order to be accurate, we identify the three landmarks on the 2-D slices. In Figure 3 (a), we find the frontozygomatic synchondrosis on the axial slice where the neighboring slices have significant morphological changes pointed by the thick arrows. The most superior posterior point of the external auditory meatus is determined as shown in Figure 3 (b). The separated skull vault and the separating plane are reconstructed in Figure 4 (b).

Skull Vault Modeling. Given the edge map $f(\mathbf{x}) : \Omega \to \mathbb{R}$ defined in $\Omega \subset \mathbb{R}^3$, the three-dimensional GVF field is defined as the vector field $\mathbf{v}(s) : \Omega \to \mathbb{R}^3$ that minimizes the energy function

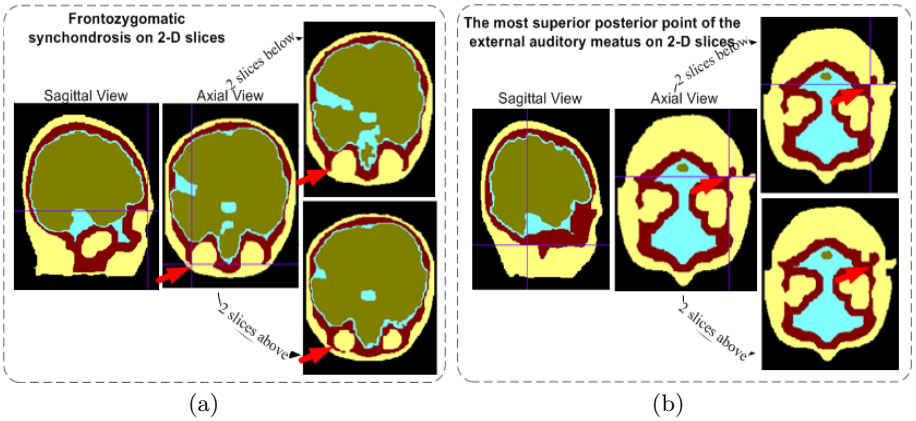
$$\epsilon = \int_{\Omega} (\mu | \bigtriangledown \mathbf{v}|^2 + | \bigtriangledown f|^2 |\mathbf{v} - \bigtriangledown f|^2) d\mathbf{s}, \tag{1}$$

where μ is the regularization parameter and f is the edge map determined using the edges on the image convoluted with a Gaussian kernel. The GVF field can be obtained by solving the Euler equations

$$\mu \bigtriangledown^2 \mathbf{v} - (\mathbf{v} - \bigtriangledown f) | \bigtriangledown f|^2 = 0, \tag{2}$$

where $\bigtriangledown^2$ is also applied to each component of the vector field $\mathbf{v}$ separately.

Distance-Based Feature Calculation. Using GVF, a smooth description of the surface is obtained. A uniformly sampled sphere (see Figure 5 (a)) is generated for each reconstructed skull vault model $m_i (i = 1, \cdots, |S|)$, where S is



(a) (b)

Fig. 3. Identification of the landmarks frontozygomatic synchondrosis, the most superior posterior point of the external auditory meatus that determine the skull vault demarcation plane. (a) Finding frontozygomatic synchondrosis. (b) Finding the most superior posterior point of the external auditory meatus.

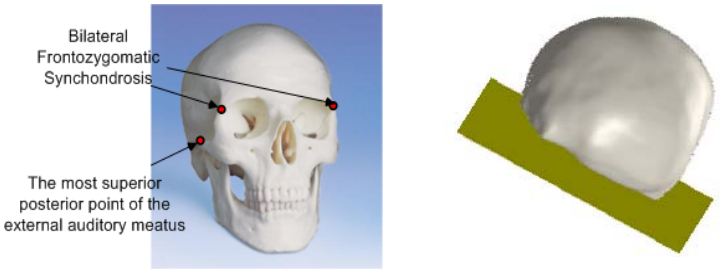


Fig. 4. Separation of the skull vault from the whole skull. (a) Positions of the three landmarks on a real skull model indicated by the red points. (b) The skull vault determined by the separating plane passing through the three anatomical landmarks.

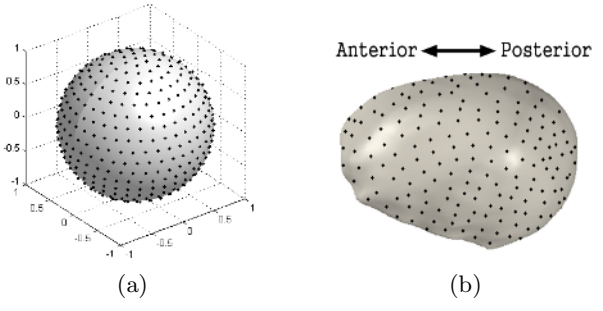


Fig. 5. Distance-based feature calculation. (a) The unit sphere centered at the skull vault centroid. (b) The spherically uniform sample points on the skull vault surface.

the set containing all subjects, and N points are sampled on each sphere. The center of sphere is translated to the centroid of m_i, and the positive direction along the y axis is rotated to the normal direction of the basal plane (or the separating plane defined in Figure 4) of m_i. Along each of these N directions, a "ray" is emitted from the center and possibly intersects with m_i. These intersections are recorded as the candidate sample points for model m_i. For each skull vault model m_i, we can find that the sample points are obtained in the same way and inherently correspond to each other. In order to ensure equal length of the resultant feature vectors, the directions that have no intersection with the surface are tagged and the corresponding sample points are removed for all subjects in S. Then the distances from the remaining sample points (see Figure 5 (b)) to the centroid of m_i are calculated as the features to describe the skull vault surface for subject m_i. These features are translationally invariant since they are all calculated relative to the centroid. And the rotational invariance is guaranteed because the basal planes have been implicitly aligned. In order for these features to be invariant to scaling, we normalize the feature vector for each subject m_i as $\mathbf{x}'_i = \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|}$.

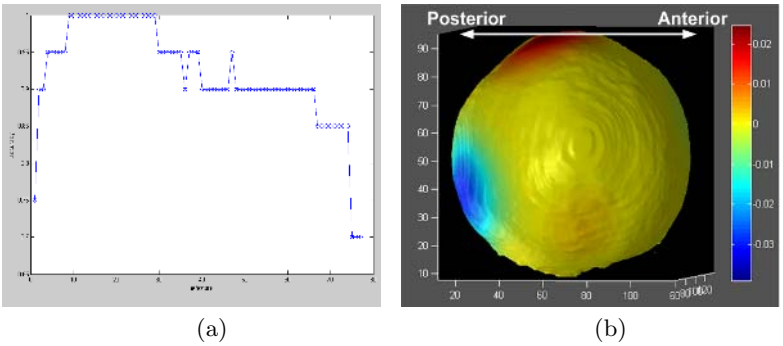
3 Pathological Abnormality Discovery Using Statistical Feature Selection

In order to select the discriminative features, we adopt a popular wrapper feature selection method, the recursive feature elimination (RFE) for support vector machine (SVM) proposed by Guyon et al. [16] to solve a cancer gene classification problem. SVM-RFE performs feature selection by iteratively training an SVM with the current set of features and removing the feature with the smallest weight in the resulting hyperplane at a time. Ultimately, this algorithm results in a ranking list of features. The number of dominating features is selected by plotting a curve of the number of features against the accuracy, and choose the number of features to be relatively small but corresponds to the highest accuracy.

4 Subjects and Results

Subjects. In the current study, the MRI data set are acquired from 10 AIS patients between 12 and 18 years old and 10 age-matched healthy girls recruited from the Scoliosis Clinic of Prince of Wales Hospital, Hong Kong. The MRI scanner used is a 1.5T MR scanner (Sonata, Siemens, Munich, Germany). The whole skull is scanned from left to right using 2-D acquisition and turbo spin-echo T2-weighted sequence with TR=6500, TE=118 ms. The thickness is 3mm without gap, and the field of view is 512×512.

Results. The effectiveness of the feature selection method is validated by the quality of the selected feature subsets in the classification. Figure 6 (a) shows the classification accuracies calculated as 1 minus the leave-one-out (LOO) errors given different numbers of selected features. One can find that the SVM-RFE selects a small feature subset which can classify AIS patients from normal controls with the highest accuracy. The classification accuracy begins to drop when too many features are added to the subset.



(a) (b)

Fig. 6. (a) The classification accuracies given different sizes of selected dominating features. (b) The group difference in skull vaults between normal controls and AIS patients mapped on the skull vault of a randomly selected normal control: top view.

The detected group difference in the skull vaults between normal controls and AIS patients are mapped on the skull vault of a randomly selected normal subject, as displayed in Figure 6 (b). The colors correspond to the values of difference: equal features for AIS and normal control are indicated by yellow; red means normal controls are statistically larger with respect to that particular feature; and blue means reversely. From the result, a direct observation is the posterior region in AIS is comparatively smaller while the left part is larger. Although it deserves further anatomical interpretations, this finding sheds light on building quantitative evidence of the ossification abnormality in AIS patients.

5 Conclusion and Future Work

This paper presents a comprehensive framework of the statistical morphology analysis for the detection of shape difference in the skull vaults between AIS patients and normal controls. To solve this particular problem, medical knowledge is combined with a variety of engineering techniques in areas of image processing, machine learning, and computer graphics. The preliminary detection result on a relatively small sample size (10 subjects in each group) is inspiring and encouraging. Based on these findings, it is promising to derive concrete evidence of systematic imbalance in the bone formation mechanisms in the AIS group.

As a pilot study, there is a lot to be explored in the future. Firstly, we plan to increase the number of subjects for more statistically reliable results. We realize that to form a clear interpretation of the current result, related findings and existing theories in other domains need to be considered. Meanwhile, the Cobb's angle of each subject, which quantifies the severity of scoliosis, can be used as the weight in statistical learning. The above work is part of an ongoing research project in the AIS Research Team using MR imaging of the whole central nervous system, which includes the skull, brain, vertebral column and spinal cord. We sought to investigate any disproportional growth between the skeletal (skull and vertebral column) and neural system (brain and spinal cord) and define the presence of neuro-osseous link in the etiopathogenesis of AIS.

Acknowledgement

The work described in this paper was supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region (Project no. CUHK4223/04E and CUHK4506/05M) and CUHK Shun Hing Institute of Advanced Engineering.

References

1. Miller, N.H.: Cause and natural history of adolescent idiopathic scoliosis. Orthop Clin North Am **30** (1999) 343–52
2. Yeung, H.: Abnormal Skeletal Growth and Bone Remodeling in Adolescent Idiopathic Scoliosis - A Morphological and Genetic Study. PhD thesis, The Chinese University of Hong Kong (2006)

3. Cheng, J., Qin, L., Cheung, C., Sher, A., Lee, K., Ng, S., Guo, X.: Generalized low areal and volumetric bone mineral density in adolescent idiopathic scoliosis. J Bone Miner Res. **15** (Aug. 2000) 1587–95
4. Cheng, J., Tang, S., Guo, X., Chan, C., Qin, L.: Osteopenia in adolescent idiopathic scoliosis: a histomorphometric study. Spine **26** (Feb. 2001) E19–23
5. Lee, W., Cheung, C., Tse, Y., Guo, X., Qin, L., Ho, S., Lau, J., Cheng, J.: Generalized low bone mass of girls with adolescent idiopathic scoliosis is related to inadequate calcium intake and weight bearing physical activity in peripubertal period. Osteoporos Int. **16** (Sep. 2005) 1024–35
6. Lee, W., Cheung, C., Tse, Y., Guo, X., Qin, L., Lam, T., Ng, B., Cheng, J.: Association of osteopenia with curve severity in adolescent idiopathic scoliosis: a study of 919 girls. Osteoporos Int. **16** (Dec. 2005) 1924–32
7. Hung, V., Qin, L., Cheung, C., Lam, T., Ng, B., Tse, Y., Guo, X., Lee, K., Cheng, J.: Osteopenia: a new prognostic factor of curve progression in adolescent idiopathic scoliosis. J Bone Joint Surg Am. **87** (Dec. 2005) 2709–16
8. Cheung, C., Lee, W., Tse, Y., Lee, K., Guo, X., Qin, L., Cheng, J.: Generalized osteopenia in adolescent idiopathic scoliosis-association with abnormal pubertal growth, bone turnover, and calcium intake? Spine **31** (Feb. 2006) 330–8
9. Cheung, C., Lee, W., Tse, Y., Tang, S., Lee, K., Guo, X., Qin, L., Cheng, J.: Abnormal peri-pubertal anthropometric measurements and growth pattern in adolescent idiopathic scoliosis: a study of 598 patients. Spine **28** (2003) 2152–2157
10. Guo, X., Chau, W., Chan, Y., Cheng, J.: Relative anterior spinal overgrowth in adolescent idiopathic scoliosis. results of disproportionate endochondral-membranous bone growth. J Bone Joint Surg Br. **85** (Sep. 2003) 1026–31
11. Guo, X., Chau, W., Chan, Y., Cheng, J., Burwell, R., Dangerfield, P.: Relative anterior spinal overgrowth in adolescent idiopathic scoliosisresult of disproportionate endochondral-membranous bone growth? summary of an electronic focus group debate of the ibse. Eur Spine J. **14** (Nov. 2005) 862–73
12. Dogdas, B., Shattuck, D.W., Leahy, R.M.: Segmentation of skull and scalp in 3-d human mri using mathematical morphology. Human Brain Mapping **26** (2005) 273–85
13. Brechbühler, C., Gerig, G., Kübler, O.: Parameterization of closed surfaces for 3-d shape description. CVGIP: Image Understanding **61** (1995) 154–70
14. Golland, P., Grimson, W., Shenton, M.E., Kikinis, R.: Small sampling size for shape analysis of anatomical structures. In: Proc MICCAI 2000, Pittsburgh, PA (2000)
15. Xu, C., Prince, J.: Snakes, shapes, and gradient vector flow. IEEE Transactions on Images Processing **7** (1998) 359–69
16. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. Journal of Machine Learning Research **3** (2003) 1157–82
17. Sandor, S., Leahy, R.: Surface-based labeling of cortical anatomy using a deformable database. IEEE Trans Med Imag **16** (1997) 41–54

A Novel Quantitative Validation of the Cortical Surface Reconstruction Algorithm Using MRI Phantom: Issues on Local Geometric Accuracy and Cortical Thickness

Junki Lee[1], Jong-Min Lee[1], Jae-Hun Kim[1], In Young Kim[1], Alan C. Evans[2], and Sun I. Kim[1]

[1] Dept. Biomedical Engineering, Hanyang University, 17 Haengdang-dong Sungdong-gu, Seoul, 133-791, Korea
[2] McConnell Brain Imaging Centre, Montreal Neurological Institute, Montreal, Canada
jklee@bme.hanyang.ac.kr

Abstract. Cortical surface reconstruction is important for functional brain mapping and morphometric analysis of the brain cortex. Several methods have been developed for the faithful reconstruction of surface models which describe the true cortical surface in both geometry and topology. However there has been no explicit method for the quantitative evaluation of the whole-cortical-surface models. In this study, we present a novel phantom-based evaluation method of the cortical surface reconstruction algorithm and quantitatively validated the local morphometric accuracy of CLASP which is one of the well-established reconstruction methods. The evaluation included local geometrical accuracy and performance of cortical thickness measure. The validation study revealed that there were some underestimations of cortical thickness measure using CLASP in the ventral and sulcal areas of the cortex and overestimations in the gyral areas and inferior temporal lobe. This study could present a generic metric for the quantitative evaluation of cortical surface reconstruction algorithm.

1 Introduction

The relationship between the cortical morphometry and the cytoarchitectonic and functional organization of the underlying cortex is a subject of much current interest and debate. Reconstructed cortices enable the visualization and the study of the sulcal and gyral patterns of an individual subject [1] and allow morphometric measurements such as cortical thickness [2]. To fulfill these needs, it is important to faithfully represent true cortical surface in terms of geometry. This task is difficult, because of artifacts such as image noise, partial volume effects, and intensity inhomogeneities [2]. Especially in tightly folded sulci, the exact boundaries of the cortex are hard to detect, because opposing sulcal banks are closer than the magnetic resonance imaging (MRI) resolution. This causes inaccuracies in surface extraction and subsequent morphometric measures such as cortical thickness.

For the evaluation of the cortical surface reconstruction, there have been various methods such as a visual validation [3], a repeatability test of the procedure [4], and a

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 183 – 190, 2006.

landmark-based validation [5]. In visual validation, the intersections of the reconstructed surface with the underlying MRI intensity data are inspected. Although this is very intuitive and easy to perform, it cannot assess the whole cortical surface in objective and quantitative way. In repeatability test, it could answer whether the procedure results in similar reconstructed surface models from different scans of the same subject. However, the accuracies of reconstructions could not be measured. In the landmark-based validation, experts select several landmark points or draw some landmark areas on the MR brain volume image. Then the closest point from each landmark to the corresponding surface is found and the distance between them computed. This distance from each landmark to the estimated cortical surface serves as a measure of accuracy. This is a quantitative validation approach. However, when experts select or draw the landmarks, some biases could be induced. In addition, it is difficult to validate the whole boundaries of cerebral cortex with this approach.

In this study, we present a novel method for the quantitative evaluation of the cortical surface reconstruction algorithm using an MRI simulator generating a realistic MRI incorporating the calculation of noise and partial volume effects. The evaluation strategy provides "gold standard" with which to access the performance of cortical surface reconstruction algorithms In previous studies, we developed a method called CLASP [3], an enhanced version of the iterative-morphing method first developed by MacDonald et al. and validated this method using a phantom-based approach. In this paper, we validated CLASP with phantom-based approach focused on the local geometric accuracy and the performance of cortical thickness measure.

2 Methods

In this phantom-based evaluation, an MRI simulator [6, 7] was used to create a "gold standard" with which to access the performance of cortical surface reconstruction algorithms directly. The accuracies of cortical geometries and thickness measures were evaluated using phantom-based comparison detailed below. T1-weighted MR images (n=12) with 1.0 mm × 1.0 mm × 1.0 mm resolution and 181 × 217 × 181 voxel dimensions were selected randomly from the datasets of the International Consortium for Brain Mapping (ICBM) [8] for this evaluation.

2.1 Cortical Surface Reconstruction and Thickness Measure

In each brain, the white (i.e. GM/WM boundary) and pial (i.e. GM/CSF boundary) surface was extracted by CLASP [3]. The CLASP algorithm consists of several stages as follows:

Acquired T1 MR images are preprocessed by intensity inhomogeneity correcti on [9] and spatial normalization to stereotaxic space [10]. Preprocessed images are classified into GM, WM, and CSF tissues [11]. The classified volumes are divided into left and right hemispheres for reconstructing two hemispheric cortical surfaces. The WM surface is reconstructed by deforming a spherical polygon model to the white matter boundary. A Laplacian field is generated between the WM surface resampled

to voxel space and a skeletonized CSF fraction image. The GM surface is initiated from the WM surface and is expanded to the boundary between GM and CSF along the Laplacian field. These stages are described in greater detail in the original paper [3]. In order to measure the cortical thickness, several preprocessing algorithms were required.

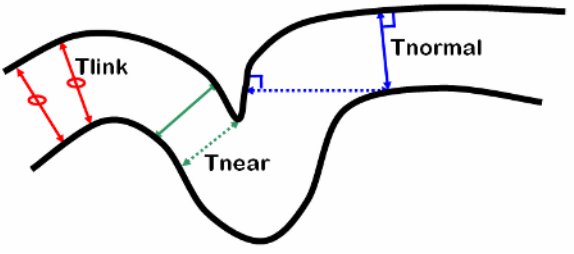


Fig. 1. Three metrics for cortical thickness measure: Tlink is the distance between corresponding points. Tnear is the distance from a vertex on the outer surface to the nearest point on the inner surface. Tnormal is the distance from a vertex on the outer surface to the nearest point on the inner surface in the direction of the surface normal. The dot lines represent improper estimations of thicknesses.

Given a surface representing the gray/CSF boundary and another representing the gray/white boundary, the cortical thickness measure could be performed [2]. Among various metrics for thickness measure, Tlink was used for the evaluation of thickness measure, which means the distance from a vertex on the outer surface to the corresponding vertex on the inner surface, as defined by the linkages used in the two-surface deformation of the CLASP algorithm (Fig. 1).The Tlink measure attempts to use the correspondence between points on the surface, which provides a measure of thickness that is less sensitive to fluctuations in surface normal and areas of high curvature. White and pial surfaces reconstructed by CLASP had the same vertex number, and the correspondence of each vertex between surfaces was defined. Thus, the cortical thickness was easily measured using the Tlink method [2]. Tnear is another approach for thickness measure which indicates the distance from a vertex on the outer surface to the nearest point on the inner surface (Fig. 1). We used Tnear metric for the measure of geometric differences between two surfaces, because it provides robust performance without the explicit correspondence of each vertex between two surfaces.

2.2 Phantom-Based Evaluation

We evaluated the cortical surface reconstruction method using a phantom-based procedure. Since there is no readily-available "gold standard" with which to assess the performance of the surface extraction algorithm, we approached the problem with an MRI simulator [6, 7]. This simulator generates a realistic MRI incorporating the calculation of noise and partial volume effects. For the validation, the following steps were performed:

1) pial and white surfaces were first extracted from the real MRI volume (Fig. 2 (a)). 2) A digital phantom including four tissue types (GM, WM, CSF, and background) was created from the surfaces. WM voxels were defined inside the white surface, and GM voxels were inserted between the pial and white surfaces. To create partial volume effects, voxels on the pial surface were given probabilities of 70% for GM and 30% for CSF. Voxels between the exterior brain mask and the pial surface were labeled as CSF. All other voxels were labeled as background (Fig. 2 (b)). 3) A T1 MR image was simulated from the phantom using the same parameters as the real data acquisition (TR=18ms, TE=10ms, slice thickness=1mm) (Fig. 2 (c)). 4) Additional substructures (skull, basal ganglia) were added from the real MRI (Fig. 2 (d)). 5) Pial and white surfaces were then extracted from the simulated MRI volume (Fig. 2 (e)). 6) Differences between each surface obtained from real or simulated MRI were measured.

In this process, the surface extracted from the real data is regarded, by definition, as "true". The experiment was designed to assess how well the surface extraction algorithm could re-capture the true surface by operating upon a simulated MRI volume derived from the true surface. CLASP was applied to the simulated MRI to generate a "test" surface. The root mean square (RMS) error between "true" and "test" surfaces then provided a measure of accuracy in surface extraction. To measure RMS error between the "true" and "test" surfaces, we calculated the distances using Tnear metric which are detailed in the previous section [2]. To validate thickness measure, we calculated the differences between "true" thicknesses and "test" thicknesses measured by Tlink. The validation of geometric accuracy and thickenss measure were performed on regional areas as well as on whole brain. To measure the local errors, we used non-rigid registration of 2-D cortical surfaces which is to find the corresponding regions of the surface model between subjects. This 2-D registration method is described in detail in next section.

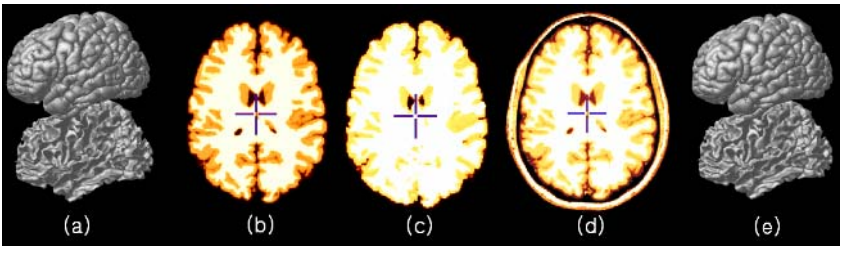


Fig. 2. Process of the evaluation using phantom (a) created cortical surface (b) surface masked volume (c) digital brain phantom (d) phantom including skull (e) recreated cortical surface from phantom

2.3 Statistical Analysis

To find the thickness of corresponding regions of the surface model between the groups, the thickness value was spatially normalized using surface based 2-D registration [12]. In the CLASP algorithm, since the cortical surfaces start from a spherical polygon model, the vertices are easily transformed to the spherical model. Vertices of each subject are nonlinearly registered to an average template on the sphere by matching crowns of gyri between subjects using a geodesic distance map [12]. Using the transformation,

thickness information and local geometric errors on the vertices were transformed to a template. Diffusion smoothing, which generalizes Gaussian kernel smoothing, with 20 mm FWHM (full-width half-maximum) was used to increase the signal-to-noise ratio of the measured cortical thickness and local geometric accuracy [13]. We then calculated mean difference between and error rate which is a difference value normalized by a thickness of "true" surface.

3 Results

The global geometric errors of the extracted cortical surfaces were measured by calculating the mean RMS distance between "true" and "test" surfaces. Fig. 3 shows mean RMS errors. The mean RMS errors was measured by Tnear (0.42 ± 0.096 mm) and Tnormal (2.73 ± 0.94 mm) respectively (Fig. 3). The local geometric errors were measured after finding correspondence between subjects. Fig. 4 shows maps of local geometric accuracies which were measured by RMS distance. RMS distance was measured by Tnear. Incorrect estimations were found in the ventral and sulcul areas of the cortex.

Cortical thicknesses using Tlink were measured and evaluated. Global/local mean differences between cortical thicknesses were measured from "true" and "test" surfaces respectively. Global mean difference of cortical thicknesses were measured on the left hemisphere (-0.061±0.24 mm), the right hemisphere (-0.080±0.20 mm), and the whole brain (-0.07±0.22 mm). Validation of local mean difference of the thickness revealed some errors in the specific areas (Fig. 5). There were some overestimations of cortical thickness measure from the "test" surface (i.e. the surface reconstructed from MRI phantom) in the ventral area and sulcus of the cortex. Underestimations were also found in the gyrus, occipital lobe, and inferior parts of temporal lobe. We measured also error rate of the cortical thickness. These values are normalized mean differences by the thicknesses of "true" surface. The overestimations of the cortical thicknesses were up to about 10% in the cingulated gyrus and some sulcal parts of the cortex. The underestimations of the cortical thicknesses were up to about 15% in the gyrus, occipital lobe, and inferior parts of temporal lobe.

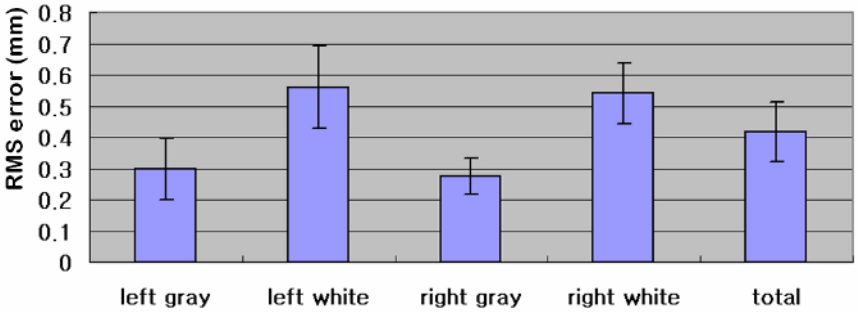


Fig. 3. Geometric errors of cortical surfaces. It was measured by calculating RMS distance between surfaces of "gold standard" and surfaces made from MR phantom images. Each column shows average RMS errors of cortical surfaces (pial/white surfaces of left/right hemispheres) created from CLASP.

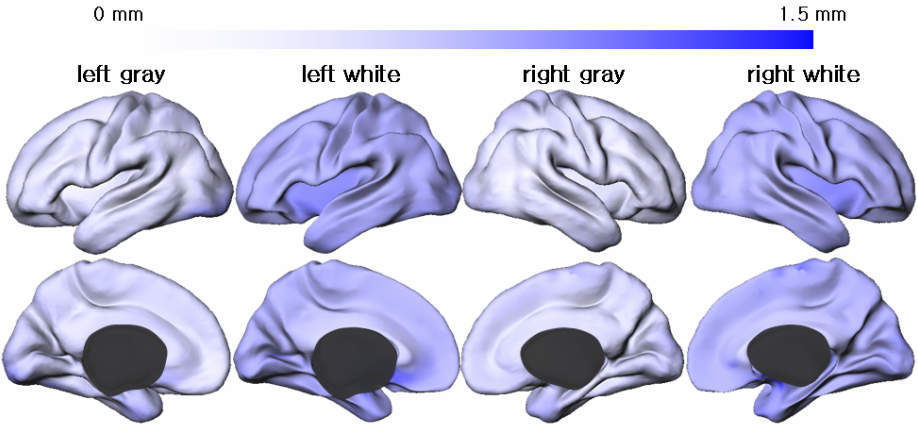


Fig. 4. Local geometric accuracies which mean the RMS distances between "true" and "test" surfaces. RMS distance was measured by Tnear. Incorrect estimations were found in the ventral and sulcul areas of the cortex.

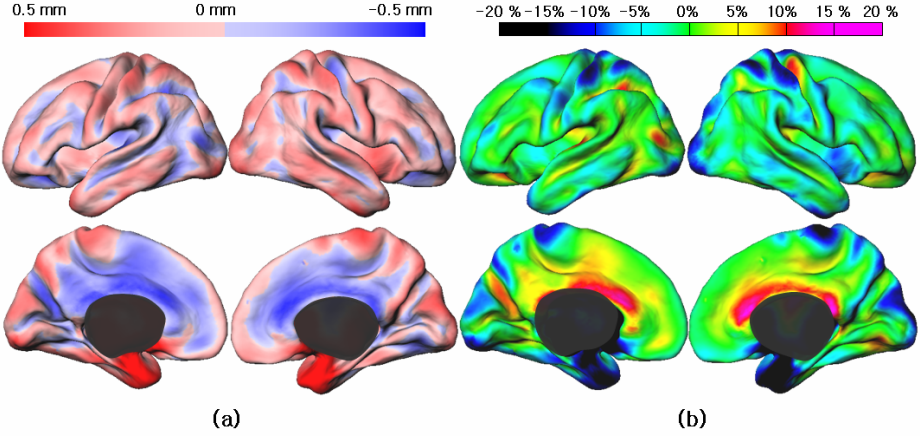


Fig. 5. (a) mean difference and (b) error rate between cortical thicknesses measured from "true" and "test" surfaces respectively. Error rate is normalized mean difference by the thicknesses of "true" surface. There were some overestimations of cortical thickness measure from the "test" surface (i.e. the surface reconstructed from MRI phantom) in the ventral area and sulcus of the cortex. The overestimations of the cortical thicknesses were up to about 10% in the cingulated gyrus and some sulcal parts of the cortex. Underestimations were also found in the gyrus, occipital lobe, and inferior parts of temporal lobe. The underestimations of the cortical thicknesses were up to about 15% in the gyrus, occipital lobe, and inferior parts of temporal lobe.

4 Discussion and Conclusions

Given the variety of cortical surface reconstruction methods, choosing an appropriate algorithm for an existing or a new problem can be quite a challenging task. Therefore,

an objective validation is necessary to provide the intrinsic characteristics of the methods, evaluate their performances and limitations. Moreover, while developing a new method, validation is essential in order to compare new and existing methods and estimate of the optimal processing parameters. However, since MR imaging, like many other medical modalities, is an in vivo study, validation becomes even more challenging and its issues are often overlooked. In this study, we conducted a quantitative validation of the performance of the cortical surface reconstruction tool using an MRI simulator. The evaluation strategy presented in this paper using MR phantom provides "gold standard" with which to access the performance of cortical surface reconstruction algorithms. We performed simulation study in which "true" surfaces, initially extracted from real MR images, were used to generate simulated MRI volumes which were then analyzed by the cortical surface reconstruction tool to recapture the original surfaces. The ability of an algorithm to re-capture this "true" surface from the simulated MRI volume, which now includes many confounding data acquisition factors (noise, loss of contrast, partial volume effects, inhomogeneity, etc.), could now be quantified. Although a simulator does not incorporate every aspect of real data, this strategy provides quantitative lower-bound performance metrics with which to assess algorithm performance.

In general, the pre-processing steps such as intensity inhomogeneity correction, skull stripping, and tissue classification are essential to the cortical surface reconstruction. However, we were mainly interested in the performance of the final surface reconstruction step out of whole reconstruction procedure including pre-processing. Therefore, the evaluation was performed using the whole procedure.

Despite of many advantages, a limitation to this phantom-based approach is that the MR phantom generated from any particular algorithm could be biased toward that method. This bias would underestimate the errors measured from the phantom-based validation. In order to overcome this limitation, it is possible that we could use another algorithm to reconstruct the phantom surfaces and cross-validate between various algorithms. The evaluation strategy presented in this paper using MR phantom provides "gold standard" with which to access the performance of cortical surface reconstruction algorithm and enables the evaluation of the specific performance which is dependent on applications such as thickness measure, surface area, fractal dimension and sulcal depth.

Acknowledgments. This work was supported by the SRC/ERC program of MOST/KOSEF (R11-2002-103).

References

[1] D. C. Van Essen, J. W. Lewis, H. A. Drury, N. Hadjikhani, R. B. Tootell, M. Bakircioglu, and M. I. Miller, "Mapping visual cortex in monkeys and humans using surface-based atlases," *Vision Res*, vol. 41, pp. 1359-78, 2001.

[2] D. MacDonald, N. Kabani, D. Avis, and A. C. Evans, "Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI," *Neuroimage*, vol. 12, pp. 340-56, 2000.

[3] J. S. Kim, V. Singh, J. K. Lee, J. Lerch, Y. Ad-Dab'bagh, D. MacDonald, J. M. Lee, S. I. Kim, and A. C. Evans, "Automated 3-D extraction and evaluation of the inner and outer cortical surfaces using a Laplacian map and partial volume effect classification," *Neuroimage*, vol. 27, pp. 210-21, 2005.
[4] N. Kriegeskorte and R. Goebel, "An efficient algorithm for topologically correct segmentation of the cortical sheet in anatomical mr volumes," *Neuroimage*, vol. 14, pp. 329-46, 2001.
[5] X. Han, D. L. Pham, D. Tosun, M. E. Rettmann, C. Xu, and J. L. Prince, "CRUISE: cortical reconstruction using implicit surface evolution," *Neuroimage*, vol. 23, pp. 997-1012, 2004.
[6] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes, and A. C. Evans, "Design and construction of a realistic digital brain phantom," *IEEE Trans Med Imaging*, vol. 17, pp. 463-8, 1998.
[7] R. K. S. Kwan, A. C. Evans, and G. B. Pike, "An extensible MRI simulator for post-processing evaluation." *Lecture Notes in Computer Science*, vol. 1131, pp. 135-140, 1996.
[8] J. C. Mazziotta, A. W. Toga, A. Evans, P. Fox, and J. Lancaster, "A probabilistic atlas of the human brain: theory and rationale for its development. The International Consortium for Brain Mapping (ICBM)," *Neuroimage*, vol. 2, pp. 89-101, 1995.
[9] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes, and A. C. Evans, "Design and construction of a realistic digital brain phantom," *IEEE Transaction on Medical Imaging*, vol. 17, pp. 463-468, 1998.
[10] D. L. Collins, P. Neelin, T. M. Peters, and A. C. Evans, "Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space," *J Comput Assist Tomogr*, vol. 18, pp. 192-205, 1994.
[11] A. P. Zijdenbos, R. Forghani, and A. C. Evans, "Automatic quantification of MS Lesions in 3D MRI brain data sets: Validation of INSECT," presented at Medical Image Computing and Computer Aided Intervention, 1998.
[12] S. Robbins, "Anatomical Standardization of the Human Brain in Euclidean 3-space and on the Cortical 2-Manifold.," *Ph.D. Thesis*, vol. School of Computer Science, McGill University, Montreal, Que., Canada, 2003.
[13] M. K. Chung, K. J. Worsley, S. Robbins, T. Paus, J. Taylor, J. N. Giedd, J. L. Rapoport, and A. C. Evans, "Deformation-based surface morphometry applied to gray matter deformation," *Neuroimage*, vol. 18, pp. 198-213, 2003.

Multivariate Statistics of the Jacobian Matrices in Tensor Based Morphometry and Their Application to HIV/AIDS

Natasha Lepore[1], Caroline A. Brun[1], Ming-Chang Chiang[1], Yi-Yu Chou[1], Rebecca A. Dutton[1], Kiralee M. Hayashi[1], Oscar L. Lopez[2], Howard J. Aizenstein[3], Arthur W. Toga[1], James T. Becker[2], and Paul M. Thompson[1]

[1] Laboratory of Neuro Imaging, Department of Neurology, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095 USA
[2] Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA 15213 USA
[3] Department of Neurology, University of Pittsburgh, Pittsburgh, PA 15213 USA*

Abstract. Tensor-based morphometry (TBM) is widely used in computational anatomy as a means to understand shape variation between structural brain images. A 3D nonlinear registration technique is typically used to align all brain images to a common neuroanatomical template, and the deformation fields are analyzed statistically to identify group differences in anatomy. However, the differences are usually computed solely from the determinants of the Jacobian matrices that are associated with the deformation fields computed by the registration procedure. Thus, much of the information contained within those matrices gets thrown out in the process. Only the magnitude of the expansions or contractions is examined, while the anisotropy and directional components of the changes are ignored.

Here we remedy this problem by computing multivariate shape change statistics using the strain matrices. As the latter do not form a vector space, means and covariances are computed on the manifold of positive-definite matrices to which they belong. We study the brain morphology of 26 HIV/AIDS patients and 14 matched healthy control subjects using our method.

The images are registered using a high-dimensional 3D fluid registration algorithm, which optimizes the Jensen-Rényi divergence, an information-theoretic measure of image correspondence. The anisotropy of the deformation is then computed. We apply a manifold version of Hotelling's T^2 test to the strain matrices. Our results complement those found from the determinants of the Jacobians alone and provide greater power in detecting group differences in brain structure.

* This research was supported by the National Institute on Aging (AG021431 and AG016570), the National Library of Medicine, the National Institute for Biomedical Imaging and Bioengineering, the National Center for Research Resources (LM05639, EB01651, RR019771), and a Research Scientist Development Award - Level II (MH01077).

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 191–198, 2006.

1 Introduction

Accurate measurement of differences in brain anatomy is key to understanding the effects of brain disorders, as well as changes associated with normal growth or variation within a population. Statistical methods that make full use of the available data can greatly improve the analysis of temporal changes within a single subject, as well as inter-subject variations. Given the subtle and distributed brain changes that occur with development and disease, improved statistical analyses are needed that detect group differences or time-dependent changes in brain structure with optimal power.

In tensor-based morphometry (TBM), a template T is matched to a study S using non-linear registration, and the displacement vector $\boldsymbol{u}(\boldsymbol{r})$ is found such that $T(\boldsymbol{r} - \boldsymbol{u})$ corresponds with $S(\boldsymbol{r})$. Here $\boldsymbol{r}$ denotes the voxel location. To help estimate anatomical correspondences, features such as point, curve, and surface landmarks present in both datasets can be used, or, more commonly, intensity-based cost functions are used based on normalized cross-correlation, mean square intensity difference, or mutual information. The Jacobian matrix of the deformation field is defined (in 3D) by

$$J = \begin{pmatrix} \partial(x-u_x)/\partial x & \partial(y-u_y)/\partial x & \partial(z-u_z)/\partial x \\ \partial(x-u_x)/\partial y & \partial(y-u_y)/\partial y & \partial(z-u_z)/\partial y \\ \partial(x-u_x)/\partial z & \partial(y-u_y)/\partial z & \partial(z-u_z)/\partial z \end{pmatrix}$$

Its determinant, the Jacobian, is most commonly used to analyze the distortion necessary to deform the images into agreement. A value $detJ(\boldsymbol{r}) > 1$ implies that the neighborhood adjacent to $\boldsymbol{r}$ in the study was stretched to match the template (i.e., local volumetric expansion), while $detJ(\boldsymbol{r}) < 1$ is associated with local shrinkage. When many subjects' images are aligned to the same standard template or atlas, maps of the Jacobians can be computed in the atlas coordinate system and group statistics can be computed at each voxel to identify localized group differences in anatomical shape or size.

However, much of the information about the shape change is lost using this measure. As a toy example of this problem, let us consider a pixel for which the eigenvalues of the Jacobian are $\lambda_{1,2,3} = \{1, 2, 0.5\}$. In such a case, the value of the Jacobian would be 1. Thus, though the eigenvalues clearly indicate directional shrinkage and growth, these changes would be left undetected, if only the Jacobian were analyzed.

In this work, we resolve this issue by making use of the strain matrices for the analysis, which are defined as $(J^T J)^{1/2}$. In particular, we apply Hotelling's T^2 test to obtain statistics for the deformation.

However, the problem is complicated by the fact that for smooth transformations, the strain matrices are constrained to the space of positive definite matrices. The latter form a cone in the space of matrices, which is itself a vector space. Thus, the strain matrices do not form a vector space, and a manifold version of the statistics is needed.

Several research groups have worked on the statistics of positive definite symmetric matrices. The first person to investigate this problem was Fréchet[1], who

defined the mean $\bar{S}$ on a manifold as the point that leads to a minimum value for the sum of squared geodesic distances between $\bar{S}$ and all the other points. $\bar{(S)}$ yields the usual value for the mean in Euclidean space, but is not necessarily unique on other manifolds. Thus, the local value of the mean is generally used instead [2], [3]. To facilitate computations, Pennec *et al.* [4], [5] and Fletcher and Joshi [6] independently proposed the use of the affine-invariant metric on the space of symmetric, positive-definite tensors. Pennec then proceeded to use the latter to define normal distributions on these matrices, as well as the Mahalanobis distance and χ^2 law. Fletcher and Joshi [6], [7] used the metric to create the principal geodesic analysis, which is an extension of principal component analysis to manifolds. The application of these techniques to calculate means of Jacobian matrices was first suggested by Woods [8]. In [9], statistics on strain tensors were used to determine the a weight factor for the regularizer in non-rigid registration. Furthermore, the Mahalanobis distance of these tensors was used as an elastic energy for the regularizer in [10]. However, to our knowledge this is the first time the full matrices are used in the context of tensor-based morphometry.

Once a metric is defined, the manifold-valued elements are projected into the tangent plane at the origin using the inverse of the exponential map, and the computations are done in this common space. Recently, Arsigny *et al.* [11], considered a new family of metrics, the 'Log-Euclidean metrics'. The latter facilitate computations, as they are chosen such that the transformed values form a vector space, and thus statistical parameters can be easily computed using standard formulae for Euclidean spaces. For two points S_1 and S_2 on the manifold, these metrics are of the form

$$d(S_1, S_2) = ||\log S_1 - \log S_2||,$$

where $||.||$ denotes a norm. Following [11], in this work we will use,

$$d(S_1, S_2) = (\text{Trace}(\log S_1 - \log S_2)^2)^{1/2}. \quad (1)$$

In this paper, we use a set of brain MRI scans from HIV/AIDS patients and matched healthy control subjects to illustrate our method. We use a fluid registration algorithm [12] to compute displacement fields and thus obtain strain matrices. All our statistics are computed within the Log-Euclidean framework. A pixelwise Hotelling T^2 test is used as a measure of variation between patients and controls. To assess the difference between our results and the ones found from the determinant of the Jacobian, these results are compared to the one-dimensional Student's t test on the determinant of the Jacobian matrices. The goal of the work was to find out whether multivariate statistics on the strain tensor afforded additional power in detecting anatomical differences between patients and controls.

The usual measure of anisotropy for tensor-valued data is the fractional anisotropy (used widely in diffusion tensor imaging) but this measure is not valid on positive definite symmetric manifolds, as it relies on a Euclidean measure for the distance. Thus, in [13], the authors propose a new measure which depends instead solely on distances computed on the manifold. They define the

geodesic anisotropy as the shortest geodesic distance between the tensor and the closest isotropic tensor within the affine-invariant metric framework. Similarly, in the Log-Euclidean metric, we can define a geodesic anisotropy GA as:

$$GA(S) = (\text{Trace}(\log S - (\langle \log S \rangle I)^2)^{1/2}, \quad \text{with } \langle \log S \rangle = \text{Trace}(\log S)/3. \quad (2)$$

2 Method

Twenty-six HIV/AIDS patients (age: 47.2 ± 9.8 years; 25M/1F; $CD4^+$ T-cell count: 299.5 ± 175.7 per μl; $\log_{10}$ viral load: 2.57 ± 1.28 RNA copies per ml of blood plasma) and 14 HIV-seronegative controls (age: 37.6 ± 12.2 years; 8M/6F) underwent 3D T1-weighted MRI scanning; subjects and scans were the same as those analyzed in the cortical thickness study by Thompson et al. [14], where more detailed neuropsychiatric data from the subjects is presented. All patients met Center for Disease Control criteria for AIDS, stage C and/or 3 (Center for Disease Control and Prevention, 1992), and none had HIV-associated dementia. Health care providers in Allegheny County, PA, served as a sentinel network for recruitment. All AIDS patients were eligible to participate, but those with a history of recent traumatic brain injury, CNS opportunistic infections, lymphoma, or stroke were excluded.

All patients underwent a detailed neurobehavioral assessment within the 4 weeks before their MRI scan, involving a neurological examination, psychosocial interview, and neuropsychological testing, and were designated as having no, mild, or moderate (coded as 0, 1, and 2 respectively) neuropsychological impairment based on a factor analysis of a broad inventory of motor and cognitive tests performed by a neuropsychologist [14].

All subjects received 3D spoiled gradient recovery (SPGR) anatomical brain MRI scans (256x256x124 matrix, TR = 25 ms, TE = 5ms; 24-cm field of view; 1.5-mm slices, zero gap; flip angle = 40^o) as part of a comprehensive neurobehavioral evaluation. The MRI brain scan of each subject was co-registered with scaling (9-parameter transformation) to the ICBM53 average brain template, after removal of extracerebral tissues (e.g., scalp, meninges, brainstem and cerebellum).

The images were non-linearly registered using a fluid registration algorithm [12], based on a convolution filter developed by Bro-Nielsen and Gramkow [15], [16] to increase the speed of the registration. As a cost function, we chose to maximize a modified version of the Jensen-Rényi Divergence (JRD). A more detailed description of our registration method can be found in [17]. To save computation time and memory requirements, the source and the target images were filtered with a Hann-windowed sinc kernel, and isotropically downsampled by a factor of 2. As in other TBM studies [18] [19], we preferred registration to a typical control image rather than a multi-subject average intensity atlas as it had sharper features. The resulting deformation field was trilinearly interpolated to drive the source image towards the target at the original resolution to obtain the warped image.

The geodesic anisotropy GA was found at each voxel using eq.2; $\tanh(GA)$ is displayed in Fig.1 a) and b). The hyperbolic tangent of the geodesic anisotropy

was used rather than GA itself, as it takes values in the interval $[0, 1]$, while those of GA span the interval $[0, \infty)$ (see [13]). We notice widespread directionality in the strain matrices, which strengthens the idea that valuable information may be found from the non-uniform components of the changes.

In order to apply Hotelling's T^2 test, we need to compute the mean and covariance matrices of the tensor-valued data. Thus, here we provide a brief summary of the method to find those quantities.

In R^n, the mean $\bar{S}$ of a set of n-dimensional vectors $S_i, i = 1, ..., m$ is the point that minimizes the summed squared distance d to all the S_i. For data on a manifold A, d becomes the geodesic distance, so $\bar{S}$ is given by [1]:

$$\bar{S} = \text{argmin}_{S \epsilon A} \Sigma_{i=1}^{m} d(S, S_i)^2. \tag{3}$$

In the log-Euclidean framework, computations are simplified by transforming the space of symmetric, positive-definite matrices into a vector space on the tangent plane, and then transforming it back to the manifold using the exponential map once the mean is taken. Thus, the formula for the mean is easily shown to be [11]:

$$\bar{S} = \exp\left(\frac{1}{M} \Sigma_{i=1}^{m} \log S_i\right) \tag{4}$$

Arsigny *et al.* demonstate that the covariance is

$$Cov(S) = \int_{\Omega} (\log(S(w)) - \log(\bar{S}(w))) \otimes (\log(S(w)) - \log(\bar{S}(w))) dP(w),$$

where dP is the probability measure, Ω is the space of possible outcomes and $w \epsilon \Omega$. For discrete data, this becomes

$$\bar{C} = \frac{1}{M} \Sigma_{i=1}^{M} \left(\log S - \log \bar{S}\right) \left(\log S - log\bar{S}\right)^T \tag{5}$$

Thus, we obtain the Mahalanobis distance

$$M = \left(\log S - log\bar{S}\right) \Sigma^{-1} \left(\log S - log\bar{S}\right)^T \tag{6}$$

So as to not assume a normal distribution, we performed a voxelwise permutation test, for which we randomly reassigned the labels of patients and controls, and compared the p-values so generated to those of the data. All our statistics were computed using 5000 permutations.

3 Results

Fig. 1 c) shows the p-values from the Student's t-test for the $\log_{10} J$ distribution. The corresponding values for Hotelling's T^2 test are shown in Fig.1 d). The p-values are considerably lower in the case of Hotelling's T^2 test. The latter is thus more sensitive, and can detect differences which are not detected with

the conventional method, even when operating on the same deformation fields. Group differences in brain structure between AIDS patients and healthy subjects are visible throughout the brain, with the greatest effect sizes in the corpus callosum and basal ganglia. That result is confirmed by prior TBM studies of HIV/AIDS [17]. The white matter also exhibits widespread atrophy, which former neuropathology studies on HIV/AIDS have shown. The cortical region is noisier, perhaps because the registration method is intensity-based and does not perform as well in that area.

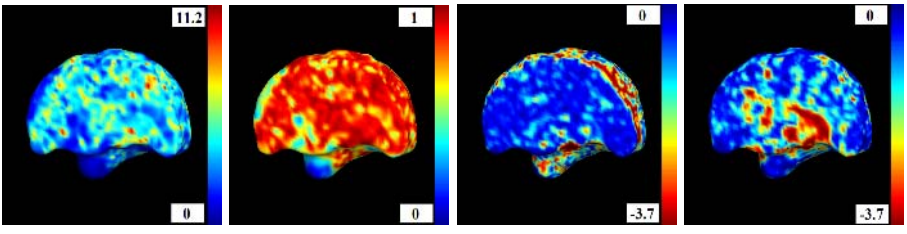


Fig. 1. From left to right: a) Voxelwise geodesic anisotropy calculated from eq.2. b) Its hyperbolic tangent c) Voxelwise p-values given by the Student's t-test for the $\log_{10}(J)$. d) Voxelwise p-values computed from the Hotelling T^2 test on the strain matrices. P-values are shown on a log_{10} scale; -3 denotes a voxel level p-value of 0.001. A sagittal section of the brain is shown, with the brain facing to the left. As expected from the literature on the neuropathology of HIV/AIDS, greatest atrophic effects are found in the subcortical regions, which border on ventricular regions enriched in the virus. The Lie group methods show comparable patterns of atrophy to the standard method, but with much greater sensitivity (i.e., better signal to noise).

4 Conclusions

Here we used a combination of tensor-based morphometry, Lie group methods, and multivariate statistics to detect systematic differences in anatomy between HIV/AIDS patients and controls. The anatomical profile of group differences is in line with studies using traditional volumetric methods, as the HIV virus is known to cause widespread neuronal loss and corresponding atrophy of the gray and white matter, especially in subcortical regions. What is more surprising is this multivariate method shows that finding with a much greater power than conventional TBM. In TBM, the Jacobian of the deformation is commonly examined, and multiple regression is applied to the scalar data from all subjects at each voxel, to identify regions of volumetric excess or deficit in one group versus another. This is clinically relevant especially given the large efforts, for example in drug trials of neurodegenerative disease, to detect subtle effects of disease on the brain using minimal numbers of subjects (or the shortest possible follow-up intervals, in a longitudinal imaging study). As such it is important to optimize not only the nonlinear registration methods that gauge alterations in brain structure, but also the statistical methods for analyzing the resulting tensor-valued

data. The current work can be expanded in several ways. First, in a multivariate setting, deformation fields could be analyzed using differential operators or tensors other than the strain tensor or deformation gradient; Rey and Cachier [20], Ashburner [21], Leow [22] and others have proposed examining the Hencky strain tensors, the logarithm of the scalar Jacobian. The optimal tensor to use for detecting group differences in TBM may depend on the directional properties of the underlying disease or developmental process being analyzed, as well as the differential operator used to regularize the deformation. Analytic formulae for the null distribution of the volume of the excursion sets of T^2-distributed random fields were recently computed by Cao and Worsley [23] [24], and these may also be applied in a Lie group setting to optimize signal detection. Studies are also underway to incorporate the log-Euclidean framework in the cost functions (or statistical priors) that regularize the nonlinear registrations themselves. This suggests that Lie group methods may be advantageous for both the registration and statistical analysis of tensor-valued data.

References

1. Fréchet, M.: Les éléments aléatoires de nature quelconque dans un espace distancié. Ann. Inst. H. Poincaré 10 (1948) 215–310
2. Karcher, H.: Riemannian center of mass and mollifier smoothing. Commun Pure Appl. Math. 30 (1977) 509–541
3. Kendall, W.S.: Probability, convexity and harmonic maps with small imageI: Uniqueness and fine existence. Proc. Lond. Math. Soc. 3 (1990) 371406
4. Pennec, X.: Probabilities and statistics on Riemannian manifolds: Basic tools for geometric measurements. In: Proceedings of the IEEE-EURASIP Workshop on Non-Linear Signal and Image Processing (NSIP99), Antalya, Turkey (1999) 194–198
5. Pennec, X., Fillard, P., Ayache, N.: A Riemannian framework for tensor computing. International Journal of Computer Vision 66 (2006) 4166 A preliminary version appeared as INRIA Research Report 5255, July 2004.
6. Fletcher, P.T., Joshi, S.: Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. In: Computer Vision ECCV 2004, Prague, Czech Republic (2004) 87–98
7. Fletcher, P.T.: Statistical Variability in Nonlinear Spaces: Application to Shape Analysis and DT-MRI. PhD thesis, Department of Computer Science, University of North Carolina, Chapel Hill, NC (2004)
8. Woods, R.P.: Characterizing volume and surface deformation in an atlas framework: theory, applications and implementation. NeuroImage 18 (2003) 769–788
9. Commowick, O., Stefanescu, R., Fillard, P., Arsigny, V., Ayache, N., Pennec, X., Malandain, G.: Incorporating statistical measures of anatomical variability in atlas-to-subject registration for conformal brain radiotherapy. In: Proceedings 8th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Part II, Palm Springs, CA, USA (2005) 927–934
10. Pennec, X., Stefanescu, R., Arsigny, V., Fillard, P., Ayache, N.: Riemannian elastic- ity: A statistical regularization framework for non-linear registration. In: Proceedings 8th International Conference on Medical Image Computing and Computer- Assisted Intervention (MICCAI), Palm Springs, CA, USA (2005) 943–950

11. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Log-Euclidean metrics for fast and simple calculus on diffusion tensors. Magnetic Resonance in Medicine (2006) To appear. Preprint: INRIA RR 5584, 2005.
12. Christensen, G., Rabbitt, R., Miller, M.: Deformable templates using large deformation kinematics. IEEE Trans. Image Process. 5 (1996) 14351447
13. Batchelor, P., Moakher, M., Atkinson, D., Calamante, F., Connelly, A.: A rigorous framework for diffusion tensor calculus. Magnetic Resonance in Medicine 53 (2005) 221–225
14. Thompson, P., Dutton, R., Hayashi, K., Toga, A., Lopez, O., Aizenstein, H., Becker, J.: Thinning of the cerebral cortex visualized in HIV/AIDS reflects CD4+T-lymphocyte decline. Proc. Natl. Acad. Sci. USA 102 (2005) 1564752
15. Bro-Nielsen, M., Gramkow, C.: Fast fluid registration of medical images. In: Proceedings of the 4th International Conference on Visualization in Biomedical Computing, Hamburg, Germany (1996) 267–276
16. Gramkow, C.: Registration of 2D and 3D medical images. Masters thesis, Danish Technical University, Copenhagen, Denmark (1996)
17. Chiang, M.C., Dutton, R.A., Hayashi, K.M., Lopez, O.L., Aizenstein, H.J., Toga, A.W., Becker, J.T., Thompson, P.M.: 3D pattern of brain atrophy in HIV/AIDS visualized using tensor-based morphometry. NeuroImage submitted (2006)
18. Studholme, C., Cardenas, V., Schuff, N., Rosen, H., Miller, B., Weiner, M.: Detecting spatially consistent structural differences in Alzheimers and fronto temporal dementia using deformation morphometry. In: Proceedings of International Conference onMedical Image Computing and Computer Assisted Intervention (MICCAI), Utrecht, Netherlands (2001) 41–48
19. Davatzikos, C., Genc, A., Xu, D., Resnick, S.: Voxel-based morphometry using the RAVENS maps: Methods and validation using simulated logitudinal atrophy. NeuroImage 14 (2001) 1361–1369
20. Cachier, P., Rey, D.: Symmetrisation of the non-rigid registration problem using inversion-invariant energies: Application to multiple sclerosis. In: Proceedings, MICCAI 2000. Volume 1935., Pittsburgh, PA (2000) 472–481
21. Ashburner, J., Friston, K.: High-dimensional image registration using symmetric priors. NeuroImage 9 (1999) 619–628
22. Leow, A.D., Geng, A., Becker, J., Davis, S.W., Toga, A.W., Thompson, P.M.: Inverse consistent mapping in 3D deformable image registration: Its construction and statistical properties, information processing in medical imaging. In: IPMI, Glenwood Springs, Colorado (2005)
23. Cao, J.: The excursion set of random fields with applications to human brain mapping. PhD thesis, Dpt. Mathematics and Statistics, McGill University, Montreal, Canada (1997)
24. Cao, J., Worsley, K.: Application of random fields in human brian mapping. In: M. Moore (Ed.) Spatial Statistics: Methodological Aspects and Applications. Volume 159. (2001) 169–182

Highly Accurate Segmentation of Brain Tissue and Subcortical Gray Matter from Newborn MRI

Neil I. Weisenfeld[1,2], Andrea U.J. Mewes[1], and Simon K. Warfield[1]

[1] Computational Radiology Laboratory
Brigham and Women's and Children's Hospitals
Harvard Medical School, Boston, MA
[2] Department of Cognitive and Neural Systems
Boston University, Boston, MA

Abstract. The segmentation of newborn brain MRI is important for assessing and directing treatment options for premature infants at risk for developmental disorders, abnormalities, or even death. Segmentation of infant brain MRI is particularly challenging when compared with the segmentation of images acquired from older children and adults. We sought to develop a fully automated segmentation strategy and present here a Bayesian approach utilizing an atlas of priors derived from previous segmentations and a new scheme for automatically selecting and iteratively refining classifier training data using the STAPLE algorithm. Results have been validated by comparison to hand-drawn segmentations.

1 Introduction

The number of preterm births in the U.S. has risen more than 30% since 1981 and the overall incidence of children born with low birthweight was 7.9% in 2003. Furthermore, "preterm birth is a leading cause of infant death and is associated with nearly one-half of all congenital neurological defects (e.g. cerebral palsy)" [1, page 17]. Regional brain tissue volume, as assessed by magnetic resonance (MR) imaging, is a biomarker for structural maturation that can be used to assess the effectiveness of interventions and for screening candidates likely to have adverse outcomes and directing them toward appropriate treatments.

MR images from the brains of newborn children are particularly difficult to segment. Aside from the occasionally unavoidable artifacts due to the motion of a non-sedated subject, neonatal brains exhibit different signal intensity characteristics than fully developed brains. The minimal brain segmentation process seeks to identify cortical gray matter with most investigators also seeking to at least identify white matter and cerebrospinal fluid. Doing so requires at least two channels of MR data acquired from different pulse sequences and it is customary to use a T1-weighted sequence and a T2-weighted sequence to differentiate CSF, gray matter, and fully-myelinated white matter.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 199–206, 2006.

In the developing brain of the normal newborn, myelination progresses in a stereotyped fashion starting with the cortico-spinal tracts. Shortly after delivery, white matter elsewhere in the brain is largely unmyelinated. Researchers studying these processes are interested in differentiating and quantifying both myelinated and unmyelinated white matter as well as cortical and sub-cortical gray matter. Differentiation of these tissue classes is not possible based solely on image intensities, so we have chosen to use an atlas of prior probabilities derived from previous segmentations. We present here a fully-automatic method for image segmentation including a new, atlas-based scheme for selection of training points for supervised classification. A number of candidate segmentations is generated for each subject, STAPLE is used to generate an optimal linear combination of candidate segmentations, and the training points for each subject are edited based on the STAPLE estimate of the true segmentation. The process is then repeated with the edited training points and iterated until a stable, final segmentation is achieved.

2 Materials and Methods

2.1 Data Acquisition

Data was acquired on a 1.5T clinical MR scanner from GE Medical Systems (Wakeusha, WI, USA) using a head-surface RF receiver coil. The present study utilizes MR scans from two pulse sequences: a T1-weighted SPGR (spoiled gradient recall) sequence and a T2-weighted scan from an FSE (fast spin echo) sequence. Additional data were acquired during these scans and not used for the present study. The scans were coronally acquired and are 256 x 256 voxels in-plane, each covering approximately 0.7mm x 0.7mm of tissue. Slice thickness was 1.5mm for the SPGR sequence and 3.0mm for the FSE sequence.

2.2 Segmentation Process

Preprocessing Preprocessing consisted of intensity non-uniformity correction, anisotropic diffusion filtering, and intra-subject rigid registration to account for subject motion between scans. For intensity non-uniformity correction we have implemented and extended the method proposed by [3] and fully developed in [4]. This method is virtually parameter-free and we have found it to be very robust. To reduce thermal noise in scans, we use a standard anisotropic diffusion [5,6] implementation provided in the open-source Insight Toolkit (ITK) [7]. Rigid registration is performed using a mutual-information implementation, also in ITK.

Atlas Construction/Registration As was mentioned previously, differentiation of sub-cortical gray matter, such a basal ganglia, and myelinated white matter from csf, unmyelinated white matter, and cortical gray matter requires information beyond image intensities alone. We have chosen to use an atlas of prior probabilities for the existence of a given tissue class at a given location (e.g. [8]). For the construction of this atlas, scans were taken from thirteen

healthy preterm infants born after 28 weeks GA and scanned at approximately 42 weeks. These scans were segmented using a previously-published semi-automated segmentation technique [8]. A model subject was chosen and scans from the remaining twelve subjects were registered to the model using an affine, mutual-information-based registration algorithm from ITK [7]. For each atlas subject, a transformation to the space of the model individual was estimated using either the T2-weighted or SPGR scan from both the model and the subject. This transformation was then applied to the individual segmentations to yield a set of aligned segmentations. Once aligned, an atlas of voxelwise probabilities was created by calculating how frequently each tissue class label appears at each voxel coordinate. In order to apply the atlas to the segmentation of a new individual, a second affine transformation is calculated to register the atlas to the new subject. This registration is performed in a similar manner as during atlas construction, only the pair-wise registration is performed using an average SPGR or T2 generated from scans of the subjects that went into the atlas itself.

Classification The goal of our segmentation process in newborns is to assign tissue class labels (background, gray matter, unmyelinated white matter, myelinated white matter, cerebrospinal fluid, and basal ganglia) to each voxel in an image volume. Even for gray matter, csf, and white matter, this is not possible based on intensities in a single image, so we consider joint intensities in both T1 and T2 weighted images and attempt to differentiate tissue classes. In order to achieve this, we generate a non-parametric estimate of the class-conditional probability density functions for each of the relevant tissue classes.

We have previously demonstrated that supervised statistical classification (classification based on the use of positive and negative examples) has advantages in terms of robustness and accuracy when compared with unsupervised classification (techniques that attempt to identify tissue classes solely from algorithmic principles) [9]. The generation of training data for supervised classification generally involves a human expert selecting a number of prototype voxels representing each tissue class and recording samples of image intensities at these locations. While this process is often not terribly time consuming, the availability of large quantities of data and larger studies makes it desirable to have a more automated method. Furthermore, manual selection of tissue class prototypes requires a trained expert, is subjective, and suffers from inter- and intra-rater variability.

We have previously [10] presented segmentation in terms of a Bayesian framework and we review the key points here and present our method for automatic selection of training points. Given a total of W tissue classes, a tissue classification can be considered the assignment of tissue class labels $w_i; i = 1..W$ to each voxel coordinate. Ultimately, we seek the posterior probability of a tissue class given the measured intensities: $p(w_i|\mathbf{x})$ where $\mathbf{x}$ is a vector of image intensity samples from each channel. Bayes' rule can be stated as $p(w_i|\mathbf{x})p(\mathbf{x}) = p(\mathbf{x}|w_i)p(w_i)$. A priori, we consider $p(\mathbf{x})$ to be uniform across intensities and it therefore represents a uniform scale factor. We estimate the posterior probability $p(w_i|\mathbf{x})$ by

first estimating the likelihood $p(\mathbf{x}|w_i)$ and then multiplying by the statical prior $p(w_i)$ provided by the atlas for each individual voxel location.

The likelihood $p(\mathbf{x}|w_i)$ is the conditional probability density function (pdf) for class w_i. The training data represents a set of samples drawn from this pdf and the process of estimating an unknown pdf from a collection of such samples is covered in detail in [11] and we summarize those results here. The probability that a sample $\mathbf{x}$ will fall in a region of the feature space R is given by $P = \int_R p(\mathbf{x}')d\mathbf{x}'$. P therefore gives a locally averaged (smoothed) estimate of $p(\mathbf{x})$ and converges to $p(\mathbf{x})$ as the region, R, gets smaller. We can estimate $p(\mathbf{x})$ therefore by estimating P. Consider a set of n samples drawn independently from $p(\mathbf{x})$ and k samples fall inside region R, the expected value for k is $E[k] = nP$ and a good estimator for P is $\frac{E[k]}{n}$. Assuming that region R is small so that $p(\mathbf{x})$ doesn't vary appreciably within the region and that $p(\mathbf{x})$ is continuous, then $P = \int_R p(\mathbf{x}')d\mathbf{x}' \approx p(\mathbf{x})V$ where V is the volume of feature space enclosed by region R. For a given trial, we assume that our k is an estimate for $E[k]$ and that an estimate for the locally averaged probability density function is $p(\mathbf{x}) = \frac{k/n}{V}$.

Furthermore, if k_i of our k samples that fall inside region R belong to class w_i, then an estimator for the joint probability is $p(\mathbf{x}, w_i) = \frac{k_i/n}{V}$. The posterior probability of any tissue class can then be estimated by [11]:

$$p(w_i|\mathbf{x})) = \frac{p(\mathbf{x}, w_i)}{\sum_{j=1}^{W} p(\mathbf{x}, w_j)} = \frac{k_i}{k}$$

This also give a robust and effective estimator for the likelihood $p(\mathbf{x}|w_i) = p(\mathbf{x}, w_i)/p(w_i)$. At this stage of analysis, the prior $p(w_i)$ is considered stationary across the image and is uniform due to a similar number of training samples in each class. The probabilities resulting from this classification are therefore a set of likelihood estimates for each class at each voxel. Simple multiplication of these likelihood estimates $p(\mathbf{x}|w_i)$ by the spatial priors encapsulated in the atlas $p(w_i)$ yields the posterior probability $p(w_i|\mathbf{x})$, since we still consider $p(\mathbf{x})$ to be uniform. A maximum aposteriori labeling of the volume is achieved by picking the class with the largest posterior value at each voxel. Prior to choosing the maximum posterior at each voxel, a Markov random field model can be utilized to provide a spatial homogeneity constraint to further refine the statistics (see, for instance, [12]).

We now turn to the problem of automatically selecting training points for a given subject. The atlas is derived from a collection of individual scans and segmentations and we'll call a subject used in the construction of the atlas an "atlas subject" and the subject being segmented the "test subject." Each atlas subject has an associated affine transformation T_a that was derived to bring the subject into alignment with a single, model subject. A second transformation T_s aligns the atlas itself to a test subject whose data is to be segmented. The composition of these two brings a coordinate from "atlas subject space" to "test subject space" and we'll call this transformation T_{as}. For each of our thirteen atlas individuals, we have a set of hand-chosen prototypes (classifier training data) that are sample locations of a given tissue class in that individual. The

prototypes were chosen to provide an effective supervised classification for that subject. A prototype is a pairing of a voxel coordinate $\mathbf{c}$ and a tissue class label w_i such as $(\mathbf{c}, w_i)$. We can transform our voxel coordinates from atlas subject space to test subject space to yield test-subject-specific prototypes $(T_{as}\{\mathbf{c}\}, w_i)$.

For each of our thirteen atlas individuals, we transform each of our prototype coordinates into the space of our new subject and use these as training points to generate a supervised classification as described above. In this manner, we generate thirteen candidate segmentations for each test subject to be segmented. The STAPLE algorithm is used to generate an optimal linear combination of these segmentations to provide a soft-labeling for each tissue class at each location in the test subject's scan. The thirteen sets of prototype points are then edited in order to reject coordinate/label pairs that the STAPLE estimate considers improbable. For each prototype point, the probability p that the true segmentation at that coordinate matches the label in the prototype is taken from the STAPLE estimate and the prototype point is either kept or rejected, accordingly. Rejection could be based on a threshold on p or, as we do here, by generation of a uniform random number between $[0, 1]$ and comparison to p. Training data that is inconsistent with the STAPLE reference standard is ultimately eliminated. This allows new estimation of the segmentations from the reduced training data, and the entire process is iterated to convergence. When a stable estimate is achieved, the STAPLE estimate can be turned into a "hard" decision by choosing the highest scoring label at each voxel location.

3 Results

Validation was performed by comparing the resulting segmentation to ground truth estimates in each of five cases. Five subjects were randomly selected from an institutional review board-approved study of neonatal brain development and none of these subjects was included in the atlas. Each subject was born premature at or before 28 weeks gestational age (GA). Two subjects exhibited no white-matter injury, two had evidence of moderate white-matter injury, and

Table 1. Results of automatic segmentation compared to single-slice manual segmentations using Dice similarity measure for cortical gray matter (cgm), myelinated white matter (myelin), unmyelinated white matter (uWM), and basal ganglia (bg). Values greater than 0.7 indicate excellent agreement.

child	cgm	csf	myelin	uWM	bg
child1	0.76	0.64	0.69	0.77	0.70
child2	0.59	0.66	0.62	0.71	0.73
child3	0.76	0.77	0.68	0.77	0.70
child4	0.77	0.63	0.77	0.74	0.64
child5	0.67	0.71	0.73	0.70	0.75
mean±sd	0.71±0.08	0.68±0.06	0.70±0.05	0.74±0.03	0.70±0.04

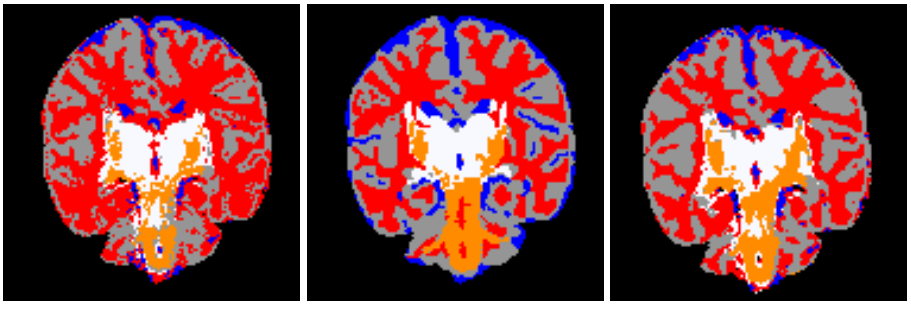


Fig. 1. Comparison of previously-published semi-automated method (left), manual segmentation (middle), and new fully-automatic method (right)

a fifth subject had severe white-matter injury. The scans used in this study were performed at term-equivalent age or approximately 42 weeks GA. These data provide a challenging test for segmentation algorithms.

In order to provide a credible estimate of ground truth, segmentations were performed by an expert rater using a combination of simple image processing operations (e.g. thresholding) and manual voxel-by-voxel editing of the results. Due to the labor intensive nature of this endeavor, a manual segmentation was generated for a single mid-coronal slice for each subject.

The segmentation quality was assessed by generation of the similarity coefficient due to Dice [13] comparing each child's automatic segmentation to the single slice manual segmentation. Table 1 shows results for the single-slice comparisons. Dice values range from zero, indicating no overlap, to one, indicating perfect overlap, but cannot be interpreted as a percentage of overlap. Zijdenbos notes that values greater than 0.7 are considered excellent agreement [14].

Quite good results were achieved for all classes including the clinically-important differentiation of myelinated and unmyelinated white matter. The worst performance was seen in the cerebrospinal fluid (CSF) tissue classes, however this is not surprising for any intensity-based classifier. In newborn images, the amount of CSF can be quite low and small regions of CSF yield image samples with large "partial-volume" effects due to the mixing of signal with that from nearby tissues.

4 Discussion and Conclusion

Automated segmentation of brain MRI has been pursued at length with promising results (e.g. [15,16,17]). Relatively few groups, however, have attempted effective segmentation algorithms for MRI of newborns. Peterson and colleagues performed a slice-by-slice manual sampling of representative points to generate tissue class averages and then thresholded and manually edited the results [18]. Prastawa and colleagues report on an automated method using a three-subject atlas for gray matter, csf, and myelinated and unmyelinated white matter, but did not attempt sub-cortical gray matter segmentation [19]. We have previously

shown a method that generates accurate segmentations, but requires interactive training [10].

Cocosco has shown [20] a method that generates prototypes for each tissue class by randomly selecting coordinates corresponding to locations where the class prior is larger than a threshold. These prototypes are then refined (pruned) using a minimum spanning tree strategy. This methodology is used in [19]. Since this method produces simplified decision boundaries, it may not fully exploit the available contrast in the images. Grau also used an atlas scheme to automatically generate prototypes, but applied this solely to adult MRI [21].

We have presented a Bayesian framework for segmentation using a supervised classifier and a strategy for automatically generating training points from an atlas of previously-generated supervised classifications. The STAPLE algorithm was used to generate an optimal linear combination of repeated segmentations in order to reduce the possibility of random chance leading to a particularly poor choice of prototype samples and to aid in the editing of samples suited to individual anatomy that, despite affine registration, still differs from the anatomy under consideration. We assessed the quality of the resulting segmentations using single-slice results drawn by an expert. Overall the strategy worked well on very challenging data. Future work should include an assessment of this technique considering the inter- and intra-rater variability of human segmentations of the same data as well as an analysis of the robustness of the method to segmentation errors in the atlas and registration failures.

Acknowledgments

This investigation was supported in part by NSF ITR 0426558, a research grant from the Whitaker Foundation, a research grant from CIMIT, grant RG 3478A2/2 from the NMSS, and by NIH grants R21 MH67054, P41 RR013218 and U41 RR019703.

References

1. Martin, J.A., Hamilton, B.E., Sutton, P.D., Ventura, S.J., Menakcker, F., Munson, M.L.: Births: Final data for 2003. National Vital Statistics Reports **54**(2) (2005)
2. Warfield, S.K., Zou, K.H., Wells, W.M.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans. Med. Imag. **23** (2004) 903–921
3. Viola, P.A.: Alignment by maximization of mutual information. PhD thesis, MIT, Cambridge, MA, USA (1995)
4. Mangin, J.F.: Entropy minimization for automatic correction of intensity non-uniformity. In: Math. Methods in Biomed. Imag. Anal., Los Alamitos, California, IEEE Computer Society (2000) 162–169
5. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Pat. Anal. Mach. Intel. **12** (1990) 629–639
6. Gerig, G., Kubler, O., Kikinis, R., Jolesz, F.: Nonlinear anistropic filtering of MRI data. IEEE Trans. Med. Imag. **2** (1992) 221–232

7. Yoo, T.S.: Insight into Images. A. K. Peters (2004)
8. Warfield, S.K., Kaus, M., Jolesz, F.A., Kikinis, R.: Adaptive, template moderated, spatially varying statistical classification. Medical Image Analysis **4**(1) (2000) 43–55
9. Warfield, S.K.: Segmentation of Magnetic Resonance Images of the Brain. PhD thesis, The University of New South Wales, Australia (1997)
10. Weisenfeld, N.I., Mewes, A.U.J., Warfield, S.K.: Segmentation of newborn brain MRI. In: Proceedings of the 2006 IEEE International Symposium on Biomedical Imaging, Piscataway, NJ, IEEE (2006) Conference held April 6-9, 2006, Crystal City, Virginia, USA.
11. Duda, R.O., Hart, P.E.: Pattern Classification and Scence Analysis. John Wiley and Sons, New York, NY, USA (1973)
12. Leemput, K.V., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based tissue classification of MR images of the brain. IEEE Trans. Med. Imag. **18** (1999) 897–908
13. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology **26** (1945) 207–302
14. Zijdenbos, A.P., Dawant, B.M., Margolin, R.A., Palmer, A.C.: Morphometric analysis of white matter lesions in MR images: Method and validation. IEEE Trans. Med. Imag. **13**(4) (1994) 716–724
15. Wells, W.M., Kikinis, R., Grimson, W.E.L., Jolesz, F.: Adaptive segmentation of MRI data. IEEE Trans. Med. Imag. **15**(4) (1996) 429–442
16. Zhang, Y., Brady, M., Smith, S.M.: Segmentation of brain MR images thorough a hidden markov random field model and the expectation maximization algorithm. IEEE Trans. Med. Imag. **20**(1) (2002) 45–57
17. Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M.: Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. Neuron **33**(3) (2002) 341–355
18. Peterson, B.S., Anderson, A.W., Ehrenkranz, R., Staib, L.H., Tageldin, M., Colson, E., Gore, J.C., Duncan, C.C., Makuch, R., Ment, L.R.: Regional brain volumes and their later neurodevelopmental correlates in term and preterm infants. Pediatrics **111** (2006) 939–948
19. Prastawa, M., Gilmore, J.H., Lin, W., Gerig, G.: Automatic segmentation of MR images of the developing newborn brain. Medical Image Analysis **9** (2005) 457–466
20. Cocosco, C.A., Zijdenbos, A.P., Evans, A.C.: A fully automatic and robust brain MRI tissue classification method. Medical Image Analysis **7** (2003) 513–527
21. Grau, V., Mewes, A.U.J., Raya, M.A., Kikinis, R., Warfield, S.K.: Improved watershed transform for medical image segmentation using prior information. IEEE Trans. Med. Imag. **23**(4) (2004) 447–458

Transformation Model and Constraints Cause Bias in Statistics on Deformation Fields

Torsten Rohlfing

Neuroscience Program, SRI International, Menlo Park, CA, USA
torsten@synapse.sri.com
http://www.stanford.edu/~rohlfing/

Abstract. This work investigates the effects of nonrigid transformation model and deformation constraints on the results of deformation-based morphometry (DBM) studies. We evaluate three popular registration algorithms: a B-spline algorithm with several different constraint terms, Thirion's demons algorithm, and a curvature PDE-based algorithm. All algorithms produced virtually identical overlaps of corresponding structures, but the underlying deformation fields were very different, and the Jacobian determinant values within homogeneous structures varied dramatically. In several cases, we observed bi-modal distributions of Jacobians within a region that violate the assumption of gaussianity that underlies many statistical tests. Our results demonstrate that, even with perfect overlap of corresponding structures, the statistics of Jacobian values are affected by bias due to design elements of the particular nonrigid registration. These findings are not limited to DBM, but also apply to voxel-based morphometry to the extent that it includes a Jacobian-based correction step ("modulation").

1 Introduction

Deformation-based morphometry (DBM) [1,2] is an increasingly popular method to assess anatomical population differences between different groups [3], or to track changes in individual subjects over time [4]. Both changes over time and differences between groups are commonly represented and analyzed as the Jacobian determinant maps of nonrigid coordinate transformations between the images involved [5]. This procedure is based on the fact that the Jacobian of the transformation at each point represents the local volume change of the coordinate space.

There is a fundamental difference between the application of nonrigid registration for the purpose of matching two images and for DBM. For multi-modality image fusion, for example, the sole criterion for success is the correct match of *visible* image features. However, when performing analysis of the deformation itself, as is done in DBM, the information stored in the deformation does not depend on the image content alone. We can, therefore, no longer rely on feature alignment. Two different deformation fields may achieve equal accuracy at all visible features, yet be very different in regions without such features. Ideally, the deformation should be equally distributed inside homogeneous structures. In the brain, for example, if there is a volume change in the ventricles, then the deformation should reflect this homogeneously rather than have substantial peaks in some areas.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 207–214, 2006.

A common view of image registration is that it is driven by two types of forces: 1) external forces that result from the dissimilarity of the two images that are being matched and that drive the transformation in the direction that improves the image match, and 2) internal forces that aim to minimize an inherent energy of the coordinate mapping, thus for example enforcing its smoothness. In homogeneous image regions, the image forces should be zero, because the image similarity measure is only sensitive to changes in the transformation that lead to a change in mapped pixel intensities. In such regions, the transformation is determined entirely by the internal forces. Our requirement as stated above thus translates into the condition that the internal forces alone should lead to a distribution of local volume changes that is maximally uniform.

On a conceptual level, a nonrigid registration method is defined by 1) an optimization method, 2) an image similarity metric, 3) a parametric coordinate transformation model, and, finally, 4) a regularizer or smoother that favors legitimate transformations over such that are not physically or otherwise reasonable. From these four building blocks, we can assume that the optimal similarity measure is determined by the types of images to be registered. The optimal registration transformation in regions with no image forces is then exclusively determined by transformation model and regularizer. We, therefore, focus in this paper on evaluating the effects that these two components have on the statistics of Jacobian determinants as they are commonly used in DBM.

2 Methods

We evaluate in this work three popular examples of nonrigid registration algorithms, for which implementations are publicly available. The particular selection of algorithms does not imply any judgment of the performance of these compared with other methods, except for our group's experience that each of the three method below is effective at successfully registering both intra-subject (i.e., longitudinal) and inter-subject images.

2.1 B-Spline Registration

Rueckert *et al.* [6] introduced a nonrigid registration algorithm that uses B-spline interpolation between uniformly spaced control points as its transformation model. The control points move independently, thus deforming the underlying image, until a local maximum of the image similarity measure is found. We use an algorithm similar to Rueckert's, but, for the purpose of this study, the sum of squared differences (SSD) image similarity measure is used rather than normalized mutual information.

The B-spline transformation model incorporates implicit smoothness and continuity constraints. Nonetheless, it was recognized early that additional constraints can be necessary to enforce physically meaningful transformations, and to incorporate prior information about permissible deformations. This is achieved by adding a weighted constraint term $E_{\mathrm{constraint}}$ (see below) to the similarity measure. We thus obtain the total cost function $E_{\mathrm{total}} = E_{\mathrm{SSD}} + \omega E_{\mathrm{constraint}}$, which is minimized during the registration. In this paper, we evaluate the following regularization approaches:

1. *Unconstrained registration*, driven only by the similarity measure, which is equivalent to a zero constraint term.

2. *Smoothness constraint* based on second-order derivatives of the transformation that models the bending energy of a thin metal plate [6,7]. For details on the computation of the constraint using derivatives of the B-spline transformation, see [8].
3. *Volume preservation constraint* as introduced by Rohlfing *et al.* [8]. The constraint penalizes the deviation of the local Jacobian from 1 (or, more generally, the global affine scaling factor) anywhere in the image. For small weights ω, the volume preservation constraint is relaxed and becomes a folding prevention constraint.

2.2 Demons Registration

Thirion [9] introduced a registration algorithm based on image forces and Gaussian regularization of the deformation field. The transformation model of this method is a local displacement field: at each and every pixel $\mathbf{x}$ the deformation is described by a displacement vector $\mathbf{u}[\mathbf{x}]$ such that $\mathbf{T}[\mathbf{x}] = \mathbf{x} + \mathbf{u}[\mathbf{x}]$. The algorithm updates the displacement field $\mathbf{u}$ iteratively. The field is updated in each step by adding a delta in the direction of the image force field to the displacement field from the previous step. For regularization, convolution with a Gaussian kernel smooths the updated deformation field after each step. The image forces for the registrations in this paper are computed by a symmetric force based on both reference and sample image gradients as suggested by [9]. We use an implementation of the algorithm that is publicly available as part of the Insight Toolkit (ITK).

2.3 Curvature-Based Registration

Variational registration algorithms casts the registration problem as a system of differential equations and solves these using variational methods. Popular examples of variational methods include elastic matching [10] and viscous fluid registration algorithms [11]. The transformation model for variational methods is a local displacement field, analogous to one used by the demons algorithm. The registration is driven by two components, a distance function D that is minimized when both images are identical, and a regularization term (also known as smoother) S that is minimized when the deformation field satisfies some smoothness constraints. The registration problem thus becomes the problem of minimizing $E_{\text{total}} = D[R, T; \mathbf{u}] + \alpha S[\mathbf{u}]$, that is, we need to minimize the image distance D while simultaneously minimizing the smoothing term S, weighted by an adjustable parameter α. In this paper in particular, we use the curvature regularization term introduced by Fischer & Modersitzki [12]. This term has several advantageous properties: it does not penalize global affine transformation components, and it can be implemented with time complexity $O(n \log n)$, where n is the number of image pixels. We have made our implementation of this algorithm publicly available by contributing it to ITK. For the purpose of this paper, the image forces that drive the registration are implemented as the symmetric demons forces as described by [9].

2.4 Electronic Phantom

We model volume gain and volume loss using a pair of simple geometric phantom images, depicted in Fig. 1. Each of the two images has 128^3 isotropic voxels and contains

(a) Inner radius 40 pixels (b) Inner radius 50 pixels (c) Triplanar slices

Fig. 1. Electronic phantom images. The pictures show the central axial slices through the isotropic 3D volume. Due to spherical symmetry, the other orthogonal central slices are identical. The radius of the outer sphere is 100 pixels in both images. The radius of the inner sphere is 40 pixels in image (a), and 50 pixels in image (b). Picture (c) shows a three-dimensional rendering of triplanar slices through the volumetric phantom image with 50 pixel inner sphere.

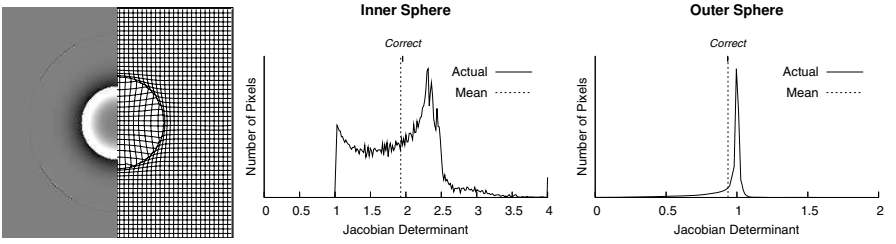
two concentric spheres. The radius of the outer sphere is 100 voxels, whereas the radius of the inner sphere is 40 voxels in one image, and 50 voxels in the other. The voxels inside the inner spheres have constant pixel intensity 50, the voxels inside the outer sphere but outside the inner sphere have constant intensity 200. The image background outside the outer sphere is set to zero. One can think of these images as extremely crude approximations of the human brain and the ventricles. Note that a key characteristic of our phantom images, due to the constant size of the outer sphere, is that they cannot be registered using a global affine transformation, i.e., we force a successful registration transformation to be truly nonrigid.

The volumes (in pixels) of the inner spheres in the two images are $V_{i,40} = 267761$ for thew sphere with radius 40 pixels, and $V_{i,50} = 523305$ for the sphere with radius 50 pixels. The volumes of the outer spheres after subtraction of the inner spheres' volumes are $V_{o,40} = 3920096$ and $V_{o,50} = 3664552$. Based on these volumes, we can compute the "ideal" Jacobian determinants J_i and J_o of the deformation field throughout the inner and outer sphere. For registration using the 40-pixel inner sphere as the reference image, these factors are $J_i = V_{i,50}/V_{i,40} \approx 1.954$ and $J_o = V_{o,50}/V_{o,40} \approx 0.935$.

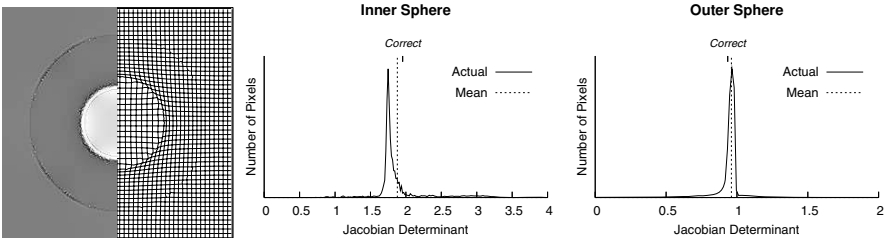
3 Results

All registration methods performed virtually indistinguishable as judged by achieving overlap between corresponding structures in the two images (recognition rates between 0.9994 and 0.9997 for all methods). The residual differences were caused by sub-pixel misalignments along the boundaries of the different structures, which was verified by visually inspecting the subtraction images.

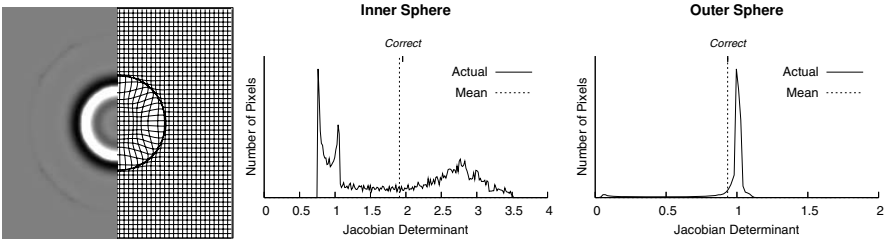
The results for the demons and curvature registration algorithms are visualized in Fig. 2. The results for the B-spline registration algorithm are visualized in Fig. 3. For each algorithm, the images on the left visualize the deformation field as gray-level encoded Jacobian values (left half) and as deformed Cartesian grids (right half). These serve to assess the spatial structure of the variation within the deformation field. The two line graphs show the distribution of Jacobians over all pixels in the inner (*left*) and outer (*right*) sphere. The average Jacobians in these graphs are represented by



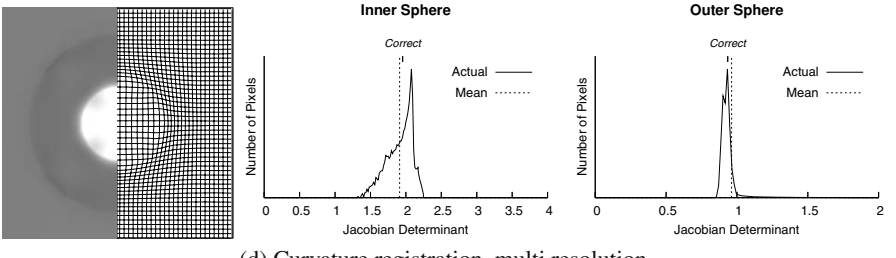
(a) Demons algorithm, single resolution



(b) Demons algorithm, multi resolution

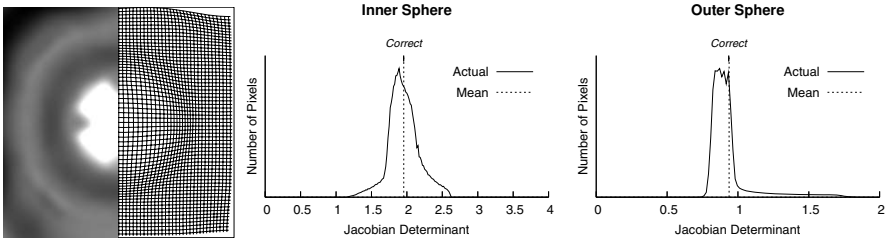


(c) Curvature registration, single resolution

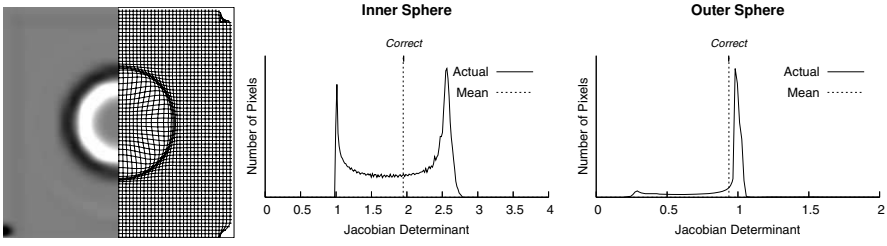


(d) Curvature registration, multi resolution

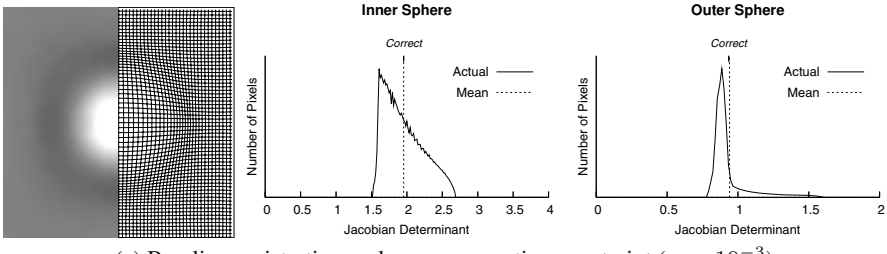
Fig. 2. Results for PDE-based registration methods. For each method we show (from left to right): Jacobian map of the coordinate transformation; deformed Cartesian grid; plots of the distributions of Jacobians for the inner and outer spheres. Note that Jacobian map and deformed grid are essentially symmetric with respect to the image center, and only one half of each is shown. Note also that the deformed grid is represented in the coordinates of the target image and thus does not match the structure of the Jacobian map, which is represented in the coordinates of the reference image.



(a) B-spline registration, unconstrained



(b) B-spline registration, smoothness constraint ($\omega = 5 \cdot 10^{-1}$)



(c) B-spline registration, volume-preservation constraint ($\omega = 10^{-3}$)

Fig. 3. Results for the B-spline registration method with different deformation constraint terms. See text and caption for Fig. 2 for details.

dashed vertical lines. Ticks labeled "Correct" mark the *a priori* correct values. All Jacobian maps are shown with a linear gray scale centered at 1.0 with a window width of 2.0.

All registration methods produced deformations fields for which the average Jacobian over each region was very close to the theoretically correct value. This is a necessary side effect of the correctly compute volume overlap. The multi-resolution demons and curvature algorithms approximated the correct value for the smaller sphere with slightly lower accuracy than the other method, but only marginally so.

The unconstrained B-spline registration (Fig. 3(a)) produced a deformation field with a noticeable geometric structure, which is determined by the properties of the B-spline. This observation, which is also illustrated in Fig. 4, previously appeared in graphical form in works by others (e.g., [5]), but was never considered in detail as a fundamental issue for subsequent deformation field analysis.

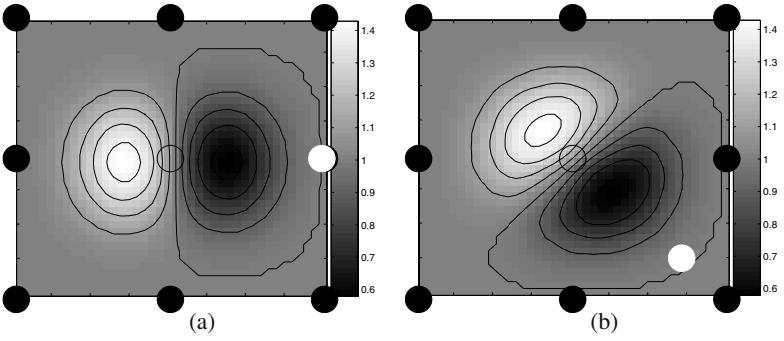


Fig. 4. Jacobian determinant map of B-spline free-form deformations with different angles of control point movement. Each image is covered by 3×3 control points (solid circles). (a) Central control point (empty circle) moved by $(+1, 0)$ grid units (to solid white circle). (b) Central control point moved by $(+\frac{1}{\sqrt{2}}, +\frac{1}{\sqrt{2}})$. Note the different shapes of the two iso-Jacobian contour sets.

Noteworthy in Fig. 2 is the improved spatial homogeneity of the Jacobian maps when a multi-resolution strategy (4 levels) is used. However, the multi-resolution strategy introduces geometrical artifacts similar to those observed in Fig. 3(a) for the unconstrained B-spline registration. Here, this effect is an artifact of refining coarser deformation fields (with larger voxels) to finer resolutions.

None of the registration algorithms produced either uniform (desirable from a conceptional point of view) or Gaussian (desirable for statistical testing) distributions of Jacobians. Several registration algorithms even produced distributions in the smaller sphere that were bi-modal: B-spline with smoothness constraint ($\omega = 5{\cdot}10^{-1}$; Fig. 3(b)), and the single-resolution demons and curvature algorithms (Fig. 2(a) and (c)). Such distributions grossly violate the assumption of Gaussian distribution of the Jacobian values that is made by most commonly applied statistical tests, e.g., t-tests.

4 Discussion

The key concept in DBM is the pixelwise comparison and statistical testing of measures derived from nonrigid coordinate transformations. The results presented in this paper demonstrate how such derived measures are determined by the absence of image features, and by the mathematical properties of the nonrigid coordinate transformation model. This includes any regularizers used to constrain the optimization problem. These effects are not random, but show strong spatial and distributional structure. DBM results may, therefore, vary greatly depending on the particular choice of the building blocks of the algorithm that is used for registration.

None of the methods tested in this work performed perfectly, although the multi-resolution demons algorithm appeared to have the relatively lowest level of artifact. Nonetheless, we found that the demons algorithm, too, produced a bi-modal, and thus highly nongaussian, Jacobian distribution when run at only a single spatial resolution. On the positive side, because for all methods the average of the Jacobian determinants

over all pixels in homogeneous regions was close to the correct volume ratios, we suggest that DBM best be performed by segmented regions rather than locally by pixels.

We note that the problems described in this paper are not limited to DBM, but also apply to application of voxel-based morphometry (VBM) [13]. The so-called "optimized VBM" [14] protocol incorporates a "modulation" step, in which the maps of brain tissue weights are corrected for inter-subject volume changes. This correction is achieved by multiplying the tissue weights at each pixel with the Jacobian determinant of the transformation that maps each subject into template space.

Acknowledgments

This work was supported by the National Institute on Alcohol Abuse and Alcoholism, Grants No. AA05965, AA10723, AA12388, AA12999, AA13521, and the National Institute on Aging, Grant No. AG17919. Many helpful comments and suggestions from A. Pfefferbaum, E.V. Sullivan, and K. Pohl are greatly appreciated.

References

1. Ashburner, J., Hutton, C., Frackowiak, *et al.*: Identifying global anatomical differences: Deformation-based morphometry. Hum. Brain Map. **6**(5–6) (1998) 348–357
2. Csernansky, J.G., Joshi, S., Wang, L., *et al.*: Hippocampal morphometry in schizophrenia by high dimensional brain mapping. Proc. Natl. Acad. Sci. USA **95**(19) (1998) 11406–11411
3. Gaser, C., Volz, H.P., Kiebel, S., *et al.*: Detecting structural changes in whole brain based on nonlinear deformations — application to schizophrenia research. NeuroImage **10**(2) (1999) 107–113
4. Shen, D., Davatzikos, C.: Very high-resolution morphometry using mass-preserving deformations and HAMMER elastic registration. NeuroImage **18**(1) (2003) 28–41
5. Chung, M.K., Worsley, K.J., Paus, T., *et al.*: A unified statistical approach to deformation-based morphometry. NeuroImage **14**(3) (2001) 595–606
6. Rueckert, D., Sonoda, L.I., Hayes, C., *et al.*: Nonrigid registration using free-form deformations: Application to breast MR images. IEEE Trans. Med. Imag. **18**(8) (1999) 712–721
7. Wahba, G.: Spline Models for Observational Data. Volume 59 of CBMS-NSF Regional Conference Series. Society for Industrial and Applied Mathematics, Philadelphia, PA (1990)
8. Rohlfing, T., Maurer, Jr., C.R., Bluemke, D.A., *et al.*: Volume-preserving nonrigid registration of MR breast images using free-form deformation with an incompressibility constraint. IEEE Trans. Med. Imag. **22**(6) (2003) 730–741
9. Thirion, J.P.: Fast non-rigid matching of 3D medical images. Technical Report 2547, INRIA, Sophia Antipolis, France (1995)
10. Bajcsy, R., Kovačič, S.: Multiresolution elastic matching. Comput. Vision. Graph. Image Processing **46**(1) (1989) 1–21
11. Christensen, G.E., Rabbitt, R.D., Miller, M.I.: Deformable templates using large deformation kinematics. IEEE Trans. Image Processing **5**(10) (1996) 1435–1447
12. Fischer, B., Modersitzki, J.: Curvature based image registration. J. Mathematical Imaging and Vision **18**(1) (2003) 81–85
13. Ashburner, J., Friston, K.J.: Voxel-based morphometry — the methods. NeuroImage **11**(6) (2000) 805–821
14. Good, C.D., Johnsrude, I.S., Ashburner, J., *et al.*: A voxel-based morphometric study of ageing in 465 normal adult human brains. NeuroImage **14**(1) (2001) 21–36

Limits on Estimating the Width of Thin Tubular Structures in 3D Images

Stefan Wörz and Karl Rohr

University of Heidelberg, IPMB, and DKFZ Heidelberg,
Dept. Bioinformatics and Functional Genomics, Biomedical Computer Vision Group
Im Neuenheimer Feld 364, 69120 Heidelberg, Germany
s.woerz@dkfz.de

Abstract. This work studies limits on estimating the width of *thin* tubular structures in 3D images. Based on nonlinear estimation theory we analyze the minimal stochastic error of estimating the width. Given a 3D analytic model of the image intensities of tubular structures, we derive a closed-form expression for the Cramér-Rao bound of the width estimate under image noise. We use the derived lower bound as a benchmark and compare it with three previously proposed accuracy limits for vessel width estimation. Moreover, by experimental investigations we demonstrate that the derived lower bound can be achieved by fitting a 3D parametric intensity model directly to the image data.

1 Introduction

Segmentation and quantification of 3D tubular structures from 3D medical image data is crucial for diagnosis, treatment, and surgical planning. In particular, the accurate quantification of vessels is indispensable. For example, an abnormal narrowing of arteries (stenosis) is one of the main reasons for heart and vascular diseases as the essential blood flow is hindered. For instance, the blocking of a coronary artery can lead to a heart attack. In clinical practice, images of the human vascular system are acquired using different imaging modalities, for example, 3D magnetic resonance angiography (MRA) or computed tomography angiography (CTA). An essential task is the accurate estimation of the width (diameter) of vessels, for example, to identify and quantify stenoses, in particular for thin vessels such as coronary arteries. For a recent review on vessel segmentation techniques we refer to [1].

Concerning *thin* vessels in 3D images, limits on the accuracy of estimating the vessel width have been studied by different groups using different approaches (e.g., [2,3,4], see below for details). However, the results of these studies have not yet been compared with each other. For a quantitative comparison of the proposed limits it would be preferable to have a benchmark. Analytic benchmarks for performance evaluation have been introduced for a different task in medical image analysis, namely, the localization of 3D landmarks in [5].

In this contribution, based on nonlinear estimation theory, we have analyzed the minimal stochastic error of estimating the width of 3D tubular structures.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 215–222, 2006.

Given a 3D analytic model of the image intensities of a tubular structure as well as white Gaussian image noise, we have derived a closed-form expression for the Cramér-Rao bound (CRB) of the vessel width, which defines the minimal uncertainty of quantifying the width. In this work, we use a cylinder model of 3D tubular structures, which is much more complex compared to the Gaussian landmark models in [5]. In addition, we consider the width of this tubular structure instead of the position of landmarks. We employ the derived CRB as a benchmark and compare it with previously proposed accuracy limits of three different approaches for vessel width estimation [2,3,4]. Moreover, by experimental investigations we demonstrate that the derived lower bound can be achieved by fitting a 3D parametric intensity model directly to the image data.

For 2D approaches addressing the accuracy of segmenting thin vessels see, for example, Sonka *et al.* [6]. Work on accuracy limits of thickness measurements of 3D *sheet-like* structures such as the vertebral cortical shell or the brain cortex has been described in [7,8], and for work on the analysis of systematic localization errors of curved surfaces see [9]. Note that the focus in these previous works is on the influence of the point spread function and sampling properties whereas the influence of image noise is not considered.

This paper is organized as follows. In the following section, we derive a closed-form expression for the CRB of a cylinder model (Sect. 2). We then present experimental results of model fitting and compare them with the theoretical results given by the CRB. In addition, we demonstrate the applicability of model fitting using different 3D medical MR images (Sect. 3). In Sect. 4, we compare the theoretical results with proposed accuracy limits of three different approaches. Finally, we give a conclusion in Sect. 5.

2 Cramér-Rao Lower Bound for 3D Cylinder Model

To derive a benchmark for performance evaluation of 3D segmentation approaches for tubular structures, we use a 3D analytic model that represents the image intensities of tubular structures. The model we employ is an ideal sharp 3D cylinder convolved with a 3D Gaussian, which is well-suited to model tubular structures of different widths (e.g., [3,4]). This 3D cylinder model comprises parameters for the width of the tubular structure (radius R), the image blur σ, as well as the image contrast a between the intensity levels of the tubular structure and the surrounding tissue. In contrast to a 3D *Gaussian* tubular model this model has two separate parameters for the vessel width and the image blur. A 2D cross-section of this Gaussian smoothed 3D cylinder is defined as

$$g_{Disk}\left(x, y, R, \sigma\right) = \mathrm{Disk}\left(x, y, R\right) \; * \; G_{\sigma}^{2D}(x, y) \tag{1}$$

where $*$ denotes convolution, $\mathrm{Disk}\left(x, y, R\right)$ is a two-valued function with value 1 if $r \leq R$ and 0 otherwise (for $r = \sqrt{x^2 + y^2}$), as well as the 2D Gaussian function $G_{\sigma}^{2D}(x, y) = G_{\sigma}(x)\, G_{\sigma}(y)$, where $G_{\sigma}(x) = \left(\sqrt{2\pi}\sigma\right)^{-1} e^{-\frac{x^2}{2\sigma^2}}$. By exploiting the

symmetries of the disk and the 2D Gaussian function as well as the separability of the 2D convolution, we can rewrite (1) as

$$g_{Disk}\,(x,y,R,\sigma) = 2\int_{-R}^{R} G_\sigma(r-\eta)\;\Phi_\sigma\Big(\sqrt{R^2-\eta^2}\Big)\;d\eta \\ -\left(\Phi_\sigma\,(r+R)-\Phi_\sigma\,(r-R)\right) \tag{2}$$

using the Gaussian error function $\Phi\,(x) = \int_{-\infty}^{x} G_1(\xi)\;d\xi$ and $\Phi_\sigma\,(x) = \Phi\,(x/\sigma)$. Extending the 2D disk in z-direction as well as including the image contrast a yields the 3D cylinder model

$$g_{M,Cylinder}(\mathbf{x},R,\sigma,a) = a\,g_{Disk}(x,y,R,\sigma)\,, \tag{3}$$

where $\mathbf{x} = (x,y,z)^T$. To determine a lower bound for estimating the radius of the tubular structure, we utilize the Fisher information matrix $\mathbf{F}$. In the case of 3D images, the Fisher information matrix is given by

$$\mathbf{F} = \frac{\mathrm{vox}^{-3}}{\sigma_n^2}\int_{-w}^{w}\int_{-w}^{w}\int_{-w}^{w}\frac{\partial g_M(\mathbf{x},\mathbf{p})}{\partial\mathbf{p}}\left(\frac{\partial g_M(\mathbf{x},\mathbf{p})}{\partial\mathbf{p}}\right)^T d\mathbf{x} \tag{4}$$

where σ_n denotes the standard deviation of the additive white Gaussian image noise, $\mathbf{p}$ are the considered model parameters, w is the half-width of the cubic region-of-interest (ROI), and *vox* denotes the spatial unit in 3D (i.e. one voxel is a cube with a size of one *vox* in each dimension). Here we are interested in estimating the radius R of the tubular structure, assuming that the values of the remaining parameters are known, i.e. $\mathbf{p} = (R)$. The matrix $\mathbf{F}$ consists of one element and the Cramér-Rao lower bound (CRB) of the uncertainty is given by (e.g., [10])

$$\sigma_{\hat{R}}^2 \geq \sigma_{CRB,\,\hat{R}}^2 = \mathbf{F}^{-1}. \tag{5}$$

The bound determines the minimal possible uncertainty of the estimated parameter R for a given level of image noise. For calculating the CRB in (5), the first order partial derivative of the cylinder model w.r.t. the radius R is required. Fortunately, whereas a closed-form solution of the Gaussian smoothed cylinder model in (3) is not known, a closed-form expression of the required partial derivative can be derived analytically:

$$\frac{\partial g_{M,Cylinder}(\mathbf{x},R,\sigma,a)}{\partial R} = a\,\frac{\sqrt{2\pi}R}{\sigma}\,G_\sigma\Big(\sqrt{r^2+R^2}\Big)\;I_0\Big(\frac{rR}{\sigma^2}\Big)\,, \tag{6}$$

with I_0 being the modified Bessel function of the first kind of order 0 (e.g., [11]).

Here, we consider cylindrical ROIs of half-width (radius) w within the xy-plane and half-width w_z in z-direction (along the cylinder) around a position on the centerline of the cylinder. A cylindrical ROI as compared to a cubic (or cuboidal) ROI is a more natural choice for tubular structures and also reduces the complexity in the calculation of the involved integrals since cylindrical coordinates can be used. Assuming that the half-width w of the ROI (within the

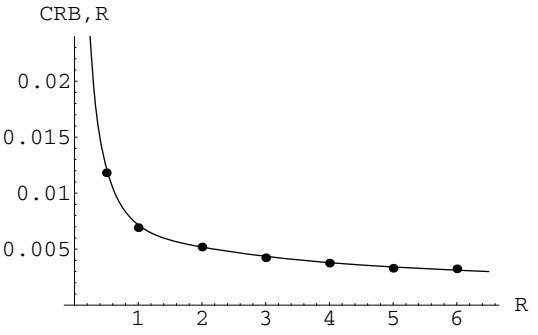


Fig. 1. Theoretical and experimental precision for estimating the radius R as a function of the radius R, using $a = 100\,\mathrm{gr}$, $\sigma = 1\,\mathrm{vox}$, $\sigma_n = 5\,\mathrm{gr}$, and $w_z = 12\,\mathrm{vox}$.

xy-plane) is much larger than the radius R and the standard deviation σ, the closed-form expression of the CRB in (5) using (6) can be stated as

$$\sigma^2_{CRB,\,\hat{R}} = \frac{\sigma^2 \sigma_n^2 \, e^{\frac{R^2}{2\sigma^2}}}{2\pi a^2 R^2 \, w_z \, I_0\left(\frac{R^2}{2\sigma^2}\right)} \,\mathrm{vox}^3. \tag{7}$$

It can be seen that the precision increases (i.e. the bound decreases) with decreasing image noise σ_n as well as increasing image contrast a and size w_z of the 3D ROI along the cylinder, and depends in a more complicated way on the radius R and the image blur σ (compared to σ_n, a, and w_z). The limit of $\sigma^2_{CRB,\,\hat{R}}$ in (7) for $R \to 0$ is ∞, and the limit for $R \to \infty$ is 0. The curve in Fig. 1 visualizes the CRB in terms of the standard deviation (square root of the variance) $\sigma_{CRB,\,\hat{R}}$ as a function of the radius R. It can be seen that the bound is monotonically decreasing as a function of the radius R.

To give an impression of the achievable accuracy, we state numerical examples of the CRB of the radius for thin tubular structures. For example, using $a = 100\,\mathrm{gr}$, $\sigma = 1\,\mathrm{vox}$, $\sigma_n = 5\,\mathrm{gr}$, and $w_z = 12\,\mathrm{vox}$, for radii R of 0.5, 1, 2, and 3 vox the minimal uncertainties $\sigma_{CRB,\,\hat{R}}$ compute to 0.012, 0.007, 0.005, and 0.004 vox, respectively (*gr* denotes the unit of the intensities in grey levels). Thus, for thin tubular structures the precision is well in the subvoxel range. However, for very thin tubular structures (e.g., a width of 1 vox) in combination with extremely poor imaging conditions (i.e. a very poor signal-to-noise ratio), the uncertainty of the radius can excess the radius of the tubular structure itself. For example, for an extremely low image contrast of $a = 5\,\mathrm{gr}$, a radius of $R = 0.5\,\mathrm{vox}$, and a small size of the ROI along the cylinder of $w_z = 5\,\mathrm{vox}$, the CRB computes to 0.76 vox, which is about 50% larger in comparison to the radius.

Note that the derived CRB in (7) does not impose a *general* limit for small radii of tubular structures (i.e. the equation is valid for $R \to 0$). However, a limit for small radii can be determined based on the desired maximal uncertainty of the radius (e.g., 5% or 0.1 vox). For example, for a maximal uncertainty of 5%, the

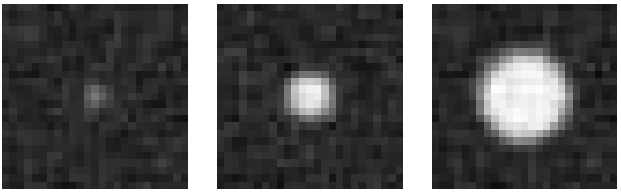


Fig. 2. 2D axial sections of 3D synthetic cylinders of radii $R = 1$ (left), $R = 3$ (middle), and $R = 6$ (right), using $a = 100\,\text{gr}$, $\sigma = 1\,\text{vox}$, $\sigma_n = 5\,\text{gr}$, and $w_z = 12\,\text{vox}$

limit computes to $R = 0.34\,\text{vox}$ (using the same parameter settings as above), i.e. the minimal width of $2R = 0.68\,\text{vox}$ is below image resolution.

3 3D Width Estimation by Model Fitting

3.1 3D Synthetic Image Data

We have carried out experiments to analyze how the theoretical bound described above relates to practice. To this end, we have generated 3D images based on the 3D cylinder model $g_{M,Cylinder}$ in (3) with additive white Gaussian noise for different radii $R = 0.5, 1, 2, 3, 4, 5, 6\,\text{vox}$ and using the same parameter settings as above (i.e. $a = 100\,\text{gr}$, $\sigma = 1\,\text{vox}$, $\sigma_n = 5\,\text{gr}$). Fig. 2 shows 2D axial sections of 3D synthetic cylinders of radii $R = 1\,\text{vox}$ (left), $R = 3\,\text{vox}$ (middle), and $R = 6\,\text{vox}$ (right). To estimate the radius of 3D tubular structures we apply a model fitting approach [4] using a cylindrical ROI of size $w = w_z = 12\,\text{vox}$. A closed-form solution of the Gaussian smoothed cylinder $g_{M,Cylinder}$ is not known. Whereas in [4] an analytic approximation has been used, we here employ numeric integration to compute the model function. Thus we have the same model for the theoretical analysis and the experiments. In total, for each value of the radius we carried out $N = 1000$ experiments (randomly varying the initial model parameters as well as the noise level) and determined the precision $\sigma_{\hat{R}}$ of the radius by the standard deviation of the estimated radii $\hat{R}_i$, i.e.

$$\sigma_{\hat{R}} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{R}_i - \bar{R}\right)^2}, \tag{8}$$

where $\bar{R}$ denotes the mean.

From the experiments it turned out that the radius estimate is unbiased, i.e. the systematic error is zero. The estimated precision $\sigma_{\hat{R}}$ of the radii are represented by the dots in Fig. 1. The curve indicates the theoretical precision according to the derived CRB $\sigma_{CRB,\hat{R}}$. It turns out that the agreement between the theoretical and the experimental values is very good, in particular, for thin vessels (e.g., $R = 0.5\,\text{vox}$), i.e. the derived lower bound can indeed be achieved experimentally. The agreement is even more remarkable since the analytic derivation does not consider discretization effects due to sampling and quantization, while in the experiments discrete images have been used.

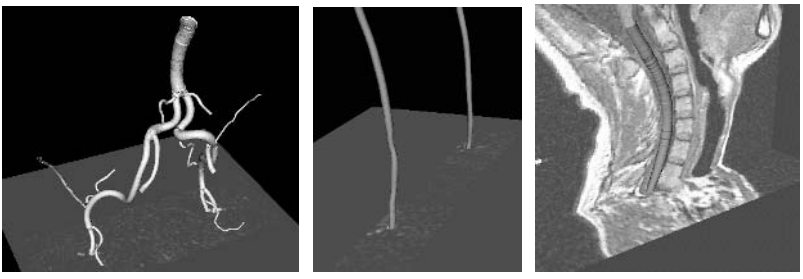


Fig. 3. Segmentation results of applying the 3D cylindrical intensity model to arteries of the pelvis (left), arteries of the thigh (middle), and to the spinal cord (right)

3.2 3D Medical MR Images

To demonstrate the applicability for real images, we have also applied the model fitting approach to 3D MR image data of the human (see Sect. 4 below for a brief description of the model fitting approach). For reasons of computational complexity we here use an analytic approximation of the Gaussian smoothed cylinder $g_{M,Cylinder}$ as in [4]. As a consequence, model fitting is about 80 to 300 times faster (depending on the size of the 3D ROI) in comparison to numerically integrating the model function. As an example, Fig. 3 shows the segmentation results for arteries of the pelvis (left) as well as for thigh arteries (middle) in 3D MRA images. It can be seen that arteries of quite different sizes including relatively thin arteries have been successfully segmented. While above we considered the segmentation of blood vessels, Fig. 3 (right) shows the segmentation result for a different type of tubular structure. Here we successfully segmented the spinal cord in a 3D MR image of the neck.

4 Comparison of Different Limits

Having derived an analytic benchmark for 3D vessel width estimation and having shown that the benchmark can be achieved in practice, we here analyze accuracy limits of three different vessel segmentation approaches and compare them with the analytic benchmark. Hoogeveen *et al.* [2] studied accuracy limits in determining the vessel width from time-of-flight (TOF) and phase-contrast (PC) 3D MRA images. Experiments were based on 3D synthetic TOF and PC MRA images as well as on real images, which were generated by using phantom tubes with known diameters. For measuring the vessel width, the criteria full-width-half-maximum and full-width-at-10% were applied for TOF and PC images, respectively. The authors state that for both TOF and PC MRA images a minimal *radius* of about *1.5 vox* is required for accurate estimation of the vessel width (allowing a maximal error of the estimated vessel width of 5%). In a second study focusing on model-based estimation of the volume flow rate from PC MRA images, Hoogeveen *et al.* [12] also provide results for estimating the vessel width. For real PC MRA images of vertebral and basilar arteries, the

authors state that their approach provides accurate vessel width estimates down to a *radius* of about *1.5 vox*, which is similar to the result in [2].

Sato *et al.* [3] developed a differential vessel segmentation approach based on a multi-scale line filter. A Hessian-based line filter is applied to different scales of a 3D image and vessels are extracted based on these filter responses. To determine the width of a vessel, the filter responses are compared to filter responses of an ideal vessel model. It turns out that a maximum response of the multi-scale filter, which is required to estimate the vessel width, is inherently not obtainable for thin vessels with a *radius* below *1.39 vox*.

In [4], we developed a model fitting approach for vessel segmentation based on an analytic 3D parametric intensity model. A 3D Gaussian smoothed cylinder is used to model the image intensities of a vessel and the surrounding tissue within a 3D ROI. Since a closed-form solution of a Gaussian smoothed cylinder is not known, we used an accurate approximation which switches between two functions, one based on the Gaussian function and the other based on the Gaussian error function. To segment a vessel we utilize an incremental process based on least-squares model fitting as well as linear Kalman filtering. The cylinder model has been applied to segment vessels in 3D MRA and 3D CTA images of the human. However, we obtain ambiguous estimates of the vessel width for a *radius* below about *1.72 vox*. The reason is that for this value we automatically switch the used approximation in our approach.

To summarize, for all three approaches [2,3,4] a limit of $R \approx 1.5$ vox has been stated. Note that, in contrast to Hoogeveen *et al.* [2], in the approach of Sato *et al.* [3] as well as in our model fitting approach [4], an accuracy limit is given by the approach itself (note also that in both approaches the limits are stated assuming a standard deviation of the Gaussian image smoothing of 1 vox). In comparison, as discussed above, the derived CRB in (7) of the vessel radius does not impose a general limit for small vessel radii such as in [3,4]. Moreover, above we have shown that the theoretically derived CRB has been achieved experimentally for thin tubular structures, in particular, for a radius of 0.5 vox. Therefore, we conclude that the limit of about $R \approx 1.5$ vox proposed by all three approaches in [2,3,4] is *not* a fundamental limit.

5 Conclusion

We have analyzed limits for estimating the width of thin 3D tubular structures such as vessels. We have derived a closed-form expression for the Cramér-Rao bound of the vessel radius and have compared this bound with three previously proposed limits. Moreover, by experimental investigations we have demonstrated that the derived lower bound can indeed be achieved by model fitting of a 3D parametric intensity model. It turns out that the previously stated limit of $R \approx 1.5$ vox is not a fundamental limit. Instead, we have shown that it is possible to correctly estimate vessels with a much smaller radius, e.g., $R = 0.5$ vox. Of course, to achieve this result certain assumptions have to be met, e.g., intensity homogeneity, straightness of the cylinder, and adherence to the Gaussian noise

model. However, all other segmentation approaches also have to make certain assumptions, and only under these assumptions the algorithms yield best results. In future work, we plan to study the performance of more complex analytic models of 3D tubular structures.

Acknowledgment

The MR images have kindly been provided by Philips Research Laboratories Hamburg, Germany as well as by Dr. med. T. Maier and Dr. C. Lienerth, Gemeinschaftspraxis Radiologie, Frankfurt/Main, Germany.

References

1. Kirbas, C., Quek, F.: A Review of Vessel Extraction Techniques and Algorithms. ACM Computing Surveys **36** (2004) 81–121
2. Hoogeveen, R., Bakker, C., Viergever, M.: Limits to the Accuracy of Vessel Diameter Measurement in MR Angiography. J. of Magnetic Resonance Imaging **8** (1998) 1228–1235
3. Sato, Y., Yamamoto, S., Tamura, S.: Accurate Quantification of Small-Diameter Tubular Structures in Isotropic CT Volume Data Based on Multiscale Line Filter Responses. In: Proc. MICCAI'04. Volume 3216 of Lecture Notes in Computer Science., St. Malo, France, Springer (2004) 508–515
4. Wörz, S., Rohr, K.: A New 3D Parametric Intensity Model for Accurate Segmentation and Quantification of Human Vessels. In: Proc. MICCAI'04. Volume 3216 of Lecture Notes in Computer Science., St. Malo, France, Springer (2004) 491–499
5. Rohr, K.: Fundamental Limits in 3D Landmark Localization. In: Proc. IPMI'05. Volume 3565 of Lecture Notes in Computer Science., Glenwood Springs, CO/USA, Springer (2005) 286–298
6. Sonka, M., Reddy, G., Winniford, M., Collins, S.: Adaptive Approach to Accurate Analysis of Small-Diameter Vessels in Cineangiograms. IEEE Trans. on Medical Imaging **16** (1997) 87–95
7. Dougherty, G., Newmann, D.: Measurement of thickness and density of thin structures by computed tomography: A simulation study. Medical Physics **26** (1999) 1341–1348
8. Sato, Y., Tanaka, H., Nishii, T., Nakanishi, K., Sugano, N., Kubota, T., Nakamura, H., Yoshikawa, H., Ochi, T., Tamura, S.: Limits on the Accuracy of 3-D Thickness Measurements in Magnetic Resonance Images – Effects of Voxel Anisotropy. IEEE Trans. on Medical Imaging **22** (2003) 1076–1088
9. Bouma, H., Vilanova, A., van Vliet, L., Gerritsen, F.: Correction for the Dislocation of Curved Surfaces Caused by the PSF in 2D and 3D CT Images. IEEE Trans. on Pattern Analysis and Machine Intelligence **27** (2005) 1501–1507
10. van Trees, H.: Detection, Estimation, and Modulation Theory, Part I. John Wiley and Sons, New York, NY/USA (1968)
11. Abramowitz, M., Stegun, I.: Pocketbook of Mathematical Functions. Verlag Harri Deutsch, Thun und Frankfurt/Main (1984)
12. Hoogeveen, R., Bakker, C., Viergever, M.: MR Phase-Contrast Flow Measurement With Limited Spatial Resolution in Small Vessels: Value of Model-Based Image Analysis. J. of Magnetic Resonance in Medicine **41** (1999) 520–528

Toward Interactive User Guiding Vessel Axis Extraction from Gray-scale Angiograms: An Optimization Framework*

Wilbur C.K. Wong and Albert C.S. Chung

Lo Kwee-Seong Medical Image Analysis Laboratory, Department of Computer Science and Engineering, The Hong Kong University of Science and Technology
{cswilbur, achung}@cse.ust.hk

Abstract. We propose a novel trace-based method to extract vessel axes from gray-scale angiograms without preliminary segmentations. Our method traces the axes on an optimization framework with the bounded spherical projection images and the sum of squared difference metric. It does not take alternate steps to search the next axial point and its tangent as in other trace-based algorithms, instead the novel method finds the solution simultaneously. This helps avoid U-turns of the trace and large spatial discontinuity of the axial points. Another advantage of the method is that it enables interactive user guidance to produce continuous tracing through regions that contain furcations, disease portions, kissing vessels (vessels in close proximity to each other) and thin vessels, which pose difficulties for the other algorithms and make re-initialization inevitable as illustrated on synthetic and clinical data sets.

1 Introduction

Segmentation of 3-D angiograms can provide patient-specific 3-D vascular models for diagnosis and prognosis. Nevertheless, if one wants to: (a) study quantitatively the vessel attributes over a particular vessel population [1]; (b) generate fly-throughs in virtual endoscopy [2]; or (c) perform real-time registration of 3-D vessels to X-ray angiograms [3], then extraction of the vessel axis is required. This usually demands further post-processing on the vascular segmentation. Topological thinning [4], voxel encoding technique [5, 2], and Voronoi diagram [6] had been investigated for the construction of vessel axes from the segmentation. Hybrid approaches were also proposed [7, 8]. In order to obtain a satisfactory vessel axis extraction with the above algorithms, a topologically and morphologically correct vascular segmentation (i.e., with no holes and cavities) is necessary, which in fact may be difficult to obtain from clinical data sets.

Skeleton-based vascular segmentation algorithms [9, 10, 11, 12] provide an alternative mean to kill two birds with one stone. In those algorithms, vessel axes are extracted and the local vessel widths are estimated in the meantime, and

* This work is supported by the HK Research Grants Council under grant 612305.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 223–231, 2006.

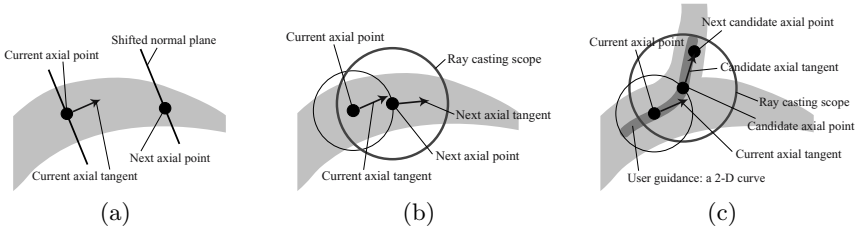


Fig. 1. Illustrations of (a) a typical trace-based method, (b) the optimization framework for axis tracing and (c) tracing with user guidance on the optimization framework

therefore a vascular segmentation can be obtained as a by-product. Technique that produces vessel axes without segmentation was also proposed [13]. Generally, those axis extraction methods [9, 11, 10, 12] consist of two core alternate tasks, the generation of the axial tangent and the computation of the next axial point. A trace is initiated by a manual seed point. The axial tangent of the seed is then generated. This tangent helps define a cross-sectional plane (normal plane), which is used to compute the next axial point after the plane is shifted forward along the tangent direction, as illustrated in Fig. 1(a). This process keeps iterating until a termination criterion is met or the user pre-empts the tracing.

Nevertheless, the trace obtained with the aforementioned methods may go astray and make U-turns in regions that contain furcations, disease portions and kissing vessels [11]. Therefore, re-initialization of the trace at the proximal end of the side branch, and the immediate disease-free and kissing-free portions is inevitable. As a result, a disconnected vessel axis may be produced.

To alleviate this problem, we propose a novel trace-based approach to extracting a continuously connected vessel axis, in which we re-formulate the axis tracing problem on an optimization framework. The objective function is built upon two bounded spherical projection images and the sum of squared difference metric. Such framework allows seamless integration of interactive user guidance (2-D curves drawn on the screen by the user) on the axis extraction so as to drive the trace along the user-defined paths to fly through those problematic regions. Our method does not take alternate steps to search the axial points and tangents as suggested in [9, 11, 10, 12]. Indeed it finds the next axial point and its tangent simultaneously and guarantees that the axial points are distanced from the preceding and the succeeding ones by a user-defined distance, which effectively helps avoid U-turns of the trace and large spatial discontinuity of the axial points. Experimental results on synthetic and clinical data sets show that our method can extract continuous and less jittering vessel axis in the regions that contain furcations, kissing vessels and thin vessels (in diameter down to 2 voxels). Furthermore, with the interactive user guidance, it can produce continuous traces to side branches at furcations and bypass disease portions.

2 Optimization Framework for Axis Tracing

2.1 Objective Function

The objective function to be optimized is based upon a bounded spherical projection (BSP) image generated at a candidate axial point with reference to its associated tangent. The BSP image is created in a similar fashion as the spherical MIP image described in [14], except the following: (1) our projection is bounded, i.e., the spherically casted rays are not extended infinitely, (2) the pixel intensities of the BSP image are the normalized cumulative sums of the voxels along the casted rays and (3) the polar coordinate is originated from the direction of the associated tangent. Figs. 2(a)-(d) show some of the BSP images generated on an image volume with a solid straight tube (intensities inside are higher than the outside) at different 3-D points and their associated tangents as depicted in Figs. 2(e)-(h). It is observed that there is a special pattern (three horizontal strips) on the BSP image if the length of the projection bound (r) is greater than the tube radius (we set r equal to 1.5× the tube radius), the 3-D point is on the desire axis and its associated tangent is aligned with the axis orientation, see Figs. 2(a) and 2(e). As such, through measuring the similarity between this BSP image ($\hat{\mathcal{I}}$) and those generated at the candidate axial points with the associated axial tangents, we can find the next axial point and the associated axial tangent simultaneously by the following minimization:

$$\boldsymbol{p}^* = \arg\min_{\boldsymbol{p}} \left(\omega_g \cdot \omega_s \cdot SSD\left(\mathcal{I}\left(\boldsymbol{p}\right), \hat{\mathcal{I}} \right) \right) \tag{1}$$

where $\boldsymbol{p}$ is a parameter vector that contains the candidate axial point, its associated tangent and the length r, $\boldsymbol{p}^*$ denotes the solution of the minimization, ω_g and ω_s are the multiplicative weights for the integration of user guidance and the constraint on axis smoothness (see Sec. 2.3 for details), the function SSD returns the sum of squared difference (SSD) metric of the two arguments, $\mathcal{I}(\boldsymbol{p})$ is the BSP image generated with the given parameter vector.

2.2 Parameter Vector

As depicted in Fig. 1(b) (which shows a 2-D version for better illustration), the location of the next axial point $\boldsymbol{x}_{i+1}$ and its associated tangent $\hat{t}_{i+1}$ can be expressed as spherical coordinates, $(\theta_x^{i+1}, \phi_x^{i+1})$ and $(\theta_t^{i+1}, \phi_t^{i+1})$, relative to the current axial points $\boldsymbol{x}_i$, respectively. The coordinates ϕ_x^{i+1} and $\phi_t^{i+1} \in [0, \pi]$ denote the polar angles originated from $\hat{t}_i$ and $(\boldsymbol{x}_{i+1} - \boldsymbol{x}_i)$, respectively, θ_x^{i+1} and $\theta_t^{i+1} \in [0, 2\pi)$ denote the azimuthal angles on the normal planes. Mathematically,

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i + \mathbf{R}_t^i \begin{bmatrix} d\cos\theta_x^{i+1} \sin\phi_x^{i+1} \\ d\sin\theta_x^{i+1} \sin\phi_x^{i+1} \\ d\cos\phi_x^{i+1} \end{bmatrix} \quad \text{and} \quad \hat{t}_{i+1} = \mathbf{R}_x^i \begin{bmatrix} \cos\theta_t^{i+1} \sin\phi_t^{i+1} \\ \sin\theta_t^{i+1} \sin\phi_t^{i+1} \\ \cos\phi_t^{i+1} \end{bmatrix}, \tag{2}$$

where $d = 1$ voxel is a user-defined step size of the trace, $\mathbf{R}_t^i$ and $\mathbf{R}_x^i$ are the 3-D rotation matrices defined by using $\hat{t}_i$ and the unit vector of $(\boldsymbol{x}_{i+1} - \boldsymbol{x}_i)$ as the Z-axis, respectively, their normal vectors as Y-axis and binormal vectors as X-axis.

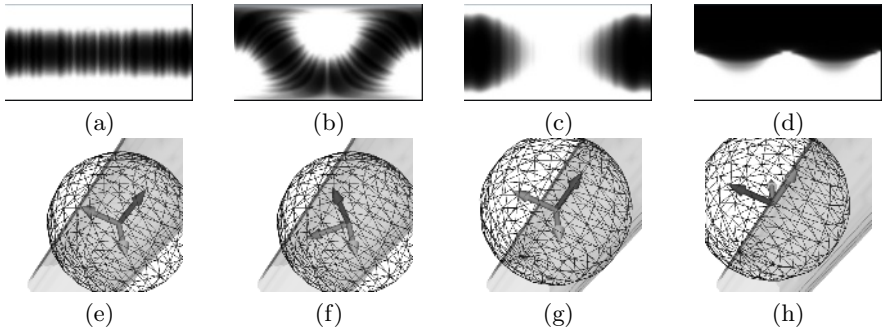


Fig. 2. (a)-(d) The BSP images (azimuthal and polar angles are enumerated in the X- and Y-axis, respectively) generated on an image volume with a solid straight tube (the surfaces are for visualization only) at different 3-D points and their associated tangents (the dark gray arrows) as depicted in (e)-(h). The wireframe-spheres define the scope of the bounded spherical projection.

Including the length of the projection bound r_{i+1} of the corresponding candidate BSP image, we have a 5-D parameter vector, $\boldsymbol{p}_{i+1} = [\theta_x^{i+1}, \phi_x^{i+1}, \theta_t^{i+1}, \phi_t^{i+1}, r_{i+1}]$. In order to avoid backward tracing, we set the search domain of the polar angles to $[0, \pi/3]$. In addition, we force $r_{i+1} \in [(1-q)r_i, (1+q)r_i]$, where $q = 0.5$ controls the allowable length change in percentage of r_i to prevent sudden changes in the length r when the trace flies through bifurcations and disease portions.

2.3 Integration of User Guidance and Axis Smoothness Constraint

User guidance is expressed as a set of user-drawn 2-D curves on the screen. They help steer the vessel axis traced in 3-D by giving reference paths on which the perspective projected trace should follow on the screen. This idea is illustrated in Fig. 1(c). Suppose the trace follows the main vessel from left to right without any guidance. If the user wants to guide the trace towards the side branch, a 2-D curve should be drawn on the screen as given.

We implement such guidance as the multiplicative weight ω_g in the objective function (see Eqn. 1). The weight is defined as follows,

$$\omega_g = \mathit{Rescale}\left(\prod_{j=1}^{2}\left(1 - \exp\left(-\frac{s_j^2}{2{\sigma_g}^2}\right)\right), \omega_g^{\min}, \omega_g^{\max}\right) \tag{3}$$

where the function *Rescale* linearly transforms the first argument from $[0, 1]$ to $[\omega_g^{\min}, \omega_g^{\max}]$, we set $\omega_g^{\min} = 0.1$ and $\omega_g^{\max} = 1$ which are the minimum and maximum values of ω_g, s_1 and s_2 are the shortest distances from the perspective projected points of the next two succeeding candidate axial points $\boldsymbol{x}'_{i+1}$ and $\boldsymbol{x}'_{i+2}$ to the 2-D curves, respectively. $\boldsymbol{x}'_{i+2}$ is estimated from the candidate axial tangent $\hat{t}'_{i+1}$ at $\boldsymbol{x}'_{i+1}$ as follows: $\boldsymbol{x}'_{i+2} = \boldsymbol{x}'_{i+1} + d\hat{t}'_{i+1}$. The variable $\sigma_g = 3$ screen voxels controls the guiding influence over the candidate axial points. A smaller weight

is obtained if $\boldsymbol{x}'_{i+1}$ and $\boldsymbol{x}'_{i+2}$ are closer to the curve on the screen. User guidance with more than one 2-D curve is demonstrated in Section 3.

Axis smoothness is constrained in the same fashion. The multiplicative weight ω_s in Eqn. 1 that favors smooth vessel axis is defined as follows,

$$\omega_s = \mathit{Rescale}\left(1 - \exp\left(-\frac{(1 - |\bar{\hat{t}} \cdot \hat{v}|)^2}{2{\sigma_s}^2}\right), \omega_s^{\min}, \omega_s^{\max}\right) \tag{4}$$

where $\omega_s^{\min} = 0.7$ and $\omega_s^{\max} = 1$ are the minimum and maximum values of the weight, $\bar{\hat{t}}$ is the mean vector of $\hat{t}_i$ and $\hat{t}'_{i+1}$, $\hat{v}$ is the normalized vector of $(\boldsymbol{x}'_{i+1} - \boldsymbol{x}_i)$ and $\sigma_s = 0.3$ controls the degree of the axis smoothness (the smaller σ_s is the smoother the axis).

2.4 Initialization and Optimization

The manually selected seed point may only be picked in close proximity to the vessel axis of interest. To initialize a trace, we need to find the first axial point $\boldsymbol{x}_0$ within the neighborhood of the seed point. This is done by locating the local maxima in a Gaussian smoothed version of the image volume. For the estimation of the initial projection bound length r_0 and the initial axial tangent $\hat{t}_0$, we perform the analysis of the Hessian matrix and compute the ridge direction (defined by $\hat{t}$) at each scale (defined by r) in a discrete scale-space [10]. Then we set the values of r_0 and $\hat{t}_0$ to the pair $(r, \hat{t})$ that gives minimum $SSD(\mathcal{I}', \hat{\mathcal{I}})$, where $\mathcal{I}'$ is the BSP image generated at $\boldsymbol{x}_0$ with the pair $(r, \hat{t})$. $\hat{t}_0$ is flipped if an opposite trace is desired. We apply a modified version of (1+1)-Evolution Strategy (ES) for the optimization (modification is referred to the optimization with simple bounds on the parameters). (1+1)-ES is a special type of the Evolutionary Algorithms with both the population size and the number of children generated equal to one, which has an automatic step size and provides search direction adaptation [15]. Thus it has the ability to step out of non-optimal minima and can provide the axis tracing problem with a long capture range.

3 Results

We have tested the proposed method on three synthetic data sets and a clinical data set (3-D rotational angiography) available at `http://www.gris.uni-tuebingen.de/areas/scivis/volren/datasets/datasets.html`, courtesy of Philips Research, Hamburg, Germany. The design of the synthetic data took several challenging vascular structures for the trace-based axis extraction methods into consideration: kissing vessels, a bifurcation, and a typical saccular aneurysm in the cerebral circulation.

Figs. 3(a)-(c) and 4 show the experimental results obtained from the newly proposed method on the synthetic and clinical data sets, respectively. The black axes in those figures are traced without any user guidance. It is evident that the vessel axes extracted from the kissing vessels (see Figs. 3(a) and 4(a)), at

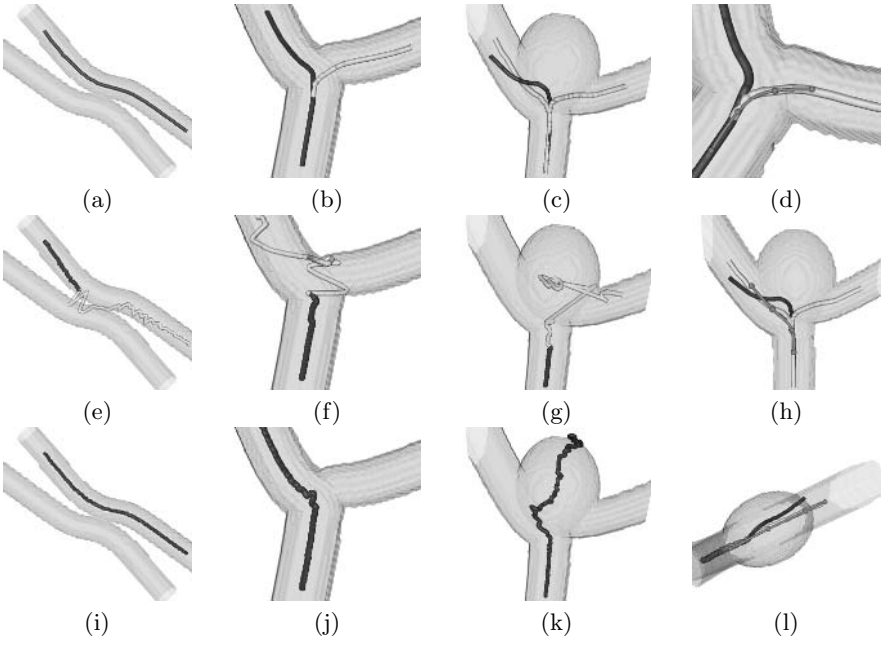
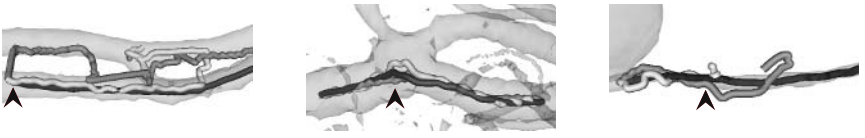


Fig. 3. Results obtained from (a)-(c) our novel method, the methods proposed by (e)-(g) Aylward and Bullitt [10] and (i)-(k) Shim *et al.* [12] on the synthetic data sets which contain kissing vessels, a bifurcation and a saccular aneurysm. The 2-D curves drawn on the screen for user guidance on the tracing are given in (d), (h) and (l). The semi-transparent surfaces are for visualization only.

the furcations (see Figs. 3(b) and 4(b)), and from the thin vessel (< 2 voxels in diameter, see Fig. 4(c)) are satisfactory. Due to the existence of the saccular aneurysm (see Fig. 3(c)), the axis extracted are steered, as a result it is deviated from the desired vessel axis. User guidance is therefore needed to drive the trace towards the target axis in this case, as well as to the side branches of the bifurcations shown in Figs. 3(b)-(c). The user-drawn 2-D curves (dark gray in color with circles, those circles are the 2-D locations picked by the user to compose the curves) are depicted in Figs. 3(d), 3(h) and 3(l). The curve shown in Fig. 3(d) helps guide the trace (light gray in color) to go to the right side branch of the bifurcation given in Fig. 3(b). The other two curves given in Figs. 3(h) and 3(l) altogether steer the trace (in light gray) to the left side branch of the saccular aneurysm shown in Fig. 3(c). Two extra 2-D curves are needed to produce the trace to the right side branch, however, they are not shown due to page limitation.

As a comparison, we have also tested the methods presented in [10, 12] on the data sets. Figs. 3(e)-(g) and 3(i)-(k) show the results obtained from Aylward and Bullitt's [10] and Shim *et al.* [12] methods on the synthetic data sets, respectively. The shorter black axes shown in Figs. 3(e)-(g) are the vessel axes extracted with the large spatial discontinuity detection (LSDD). Such



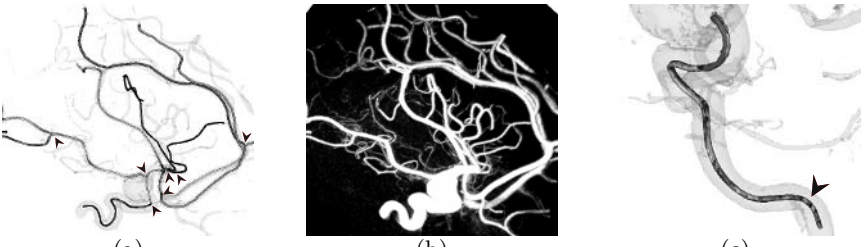
(a) (b) (c)

Fig. 4. Results on the clinical data set. The black, dark and light gray axes are the vessel axes obtained from ours, Aylward and Bullitt's [10] and Shim *et al.* [12] methods, respectively, on different vascular regions. The semi-transparent surfaces are for visualization only.

detection mechanism preempts the tracing if the next axial points is > 1 voxel apart from the current one. It is observed that, with the detection, Aylward and Bullitt's method stops the tracing just before entering the kissing region, bifurcation and disease region. The method, however, produces axes (in gray) with a large axial point displacement in the aforementioned regions if no LSDD. Shim *et al.* method gives similar results except for the kissing vessels. In addition, several U-turns can be observed at the bifurcation and in the aneurysmal sac. Comparisons on the clinical data set are presented in Fig. 4, in which the dark and light gray axes are the vessel axes obtained from Aylward and Bullitt's, and Shim *et al.* methods, respectively. In the regions that contain kissing vessels, a furcation and a thin vessel, the axes extracted from the two methods is either jittered in between the two adjacent vessels, dragged towards the center of the furcation or off the target axis when the vessel is too thin to be traced. In general, the compared two methods produce vessel axis with more jitters than ours.

Furthermore, we have extracted the axes of nine major vessels, including the internal carotid artery (ICA), the two anterior cerebral arteries (ACA), a few branches of the middle cerebral artery (MCA) and the posterior communicating artery (PCoA), in the clinical data set with our method. They are depicted in Fig. 5(a). An MIP image is shown in Fig. 5(b) for reference. User guidance is given at the furcations to steer the traces towards the side branches and at the base of the aneurysm to bypass the disease portion (highlighted by arrows). It is demonstrated that our method can trace vessel segments, including the side branches and that in the disease region. Since the other algorithms do not support user guidance, they are not tested in this study. To study the repeatability of the method, the ICA is traced for five times. The axes extracted are shown in Fig. 5(c). Visually, there is no significant discrepancy amongst those axes.

We have also evaluated quantitatively the extracted vessel axes with the manually drawn axes built with reference to the reformatted cross-sectional images along the vessels. The mean Euclidean distance (ED) errors (maximum errors are given in the parentheses) in percentage of diagonal length of the voxel (ΔV) are listed in Table 1. The error is defined as the shortest ED from an axial point of the extracted axis to the manually drawn axis. Apparently Aylward and Bullitt's method with LSDD performs better according to the errors tabulated. However, it produces preempted traces upon large spatial discontinuity is encountered (the locations are highlighted by arrows in Figs. 4 and 5(c)). Compared with



(a) (b) (c)

Fig. 5. (a) Extracted axes of nine major vessels in the clinical data set, including ICA, two ACA, a few branches of MCA and PCoA. (b) MIP image. (c) Axis of ICA extracted at five difference times. The semi-transparent surfaces are for visualization only.

Table 1. The Euclidean distance errors in percentage of diagonal length of the voxel between the extracted vessel axes and the manually drawn axes. The mean errors are tabulated with the maximum errors in the parentheses. The synthetic data sets are those shown in Fig. 3. The clinical data sets of sub-regions, major vessels and ICA are those presented in Figs. 4, 5(a) and 5(c), respectively.

Data Sets	Ours	Aylward and Bullitt with LSDD [10]	Aylward and Bullitt [10]	Shim *et al.* [12]
Synthetic	29% (126%)	16% (140%)	464% (1274%)	304% (798%)
Clinical (Sub-regions)	23% (60%)	10% (53%)	96% (182%)	70% (228%)
Clinical (Major vessels)	29% (100%)	—	—	—
Clinical (ICA)	32% (77%)	3% (76%)	46% (137%)	6% (86%)

the other algorithms which produce axes with large errors in the synthetic and several sub-regions of the clinical data sets, our method is capable of extracting the complete axes of the vessel segments consistently with mean errors $< \Delta V/3$ and maximum errors $\approx \Delta V$. Concerning the processing time, the axes of the nine major vessels took < 15 min to extract with the novel method on a 2.0 GHz PC. The manual process, by contrast, took 169 min to finish.

4 Discussion and Conclusion

We have proposed a novel trace-based method which enables interactive user guiding of the trace to fly through challenging areas, e.g., vascular regions that contain furcations, disease portions, kissing vessels and thin vessels. The method traces a vessel axis on an optimization framework with the bounded spherical projection (BSP) images and the sum of squared difference (SSD) metric. Interactive user guidance (2-D curves that drawn on the screen by the user to steer the trace) and axis smoothness constraint are integrated seamlessly as the multiplicative weights in the objective function. Experimental results on synthetic and clinical data have shown that our method can extract continuous and less jittering vessel axes, as compared with the other two trace-based algorithms.

Several potential extensions to the method are of interest for further research. For instance, reducing the user guidance on tracing the side branches of a furcation, extending the method to trace axes of anatomies other than blood vessels (especially those with asymmetric cross-sections) with different reference images/models and similarity measures (i.e. other than BSP images and SSD metric), and incorporating the skeleton-based segmentation algorithms so as to help producing topologically and morphologically correct vascular segmentations.

References

1. Bullitt, E., et al.: Analyzing attributes of vessel populations. MedIA **9** (2004) 39–49
2. Farag, A.A., et al.: Reliable fly-throughs of vascular trees. Technical report, Dept. Elec. Comput. Eng., Univ. of Louisville (2004)
3. Kita, Y., et al.: Real-time registration of 3D cerebral vessels to X-ray angiograms. In: MICCAI. (1998) 1125–1133
4. Palágyi, K., et al.: A sequential 3D thinning algorithm and its medical applications. In: IPMI. (2001) 409–415
5. Zhou, Y., Toga, A.W.: Efficient skeletonization of volumetric objects. TVCG **5** (1999) 196–209
6. Ogniewicz, R.L., Kubler, O.: Hierarchy Voronoi skeletons. PR **28** (1995) 343–359
7. Flasque, N., et al.: Acquisition, segmentation and tracking of the cerebral vascular tree on 3D magnetic resonance angiography images. MedIA **5** (2001) 173–183
8. Hassouna, M.S., et al.: Differential fly-throughs (DFT): a general framework for computing flight paths. In: MICCAI. (2005) 654–661
9. Hernández-Hoyos, M., et al.: A deformable vessel model with single point initialization for segmentation, quantification and visualization of blood vessels in 3D MRA. In: MICCAI. (2000) 735–745
10. Aylward, S., Bullitt, E.: Initialization, noise, singularities, and scale in height ridge traversal for tubular object centerline extraction. TMI **21** (2002) 61–75
11. Wink, O., et al.: Fast delineation and visualization of vessels in 3-D angiographic images. TMI **19** (2000) 337–346
12. Shim, H., et al.: Partition-based extraction of cerebral arteries from CT angiography with emphasis on adaptive tracking. In: IPMI. (2005) 357–368
13. Krissian, K., et al.: Fast and automatic vessel centerline detection for MRA. In: ISMRM. (2002)
14. van der Weide, R., et al.: CTA-based angle selection for diagnostic and interventional angiography of saccular intracranial aneurysms. TMI **17** (1998) 831–841
15. Styner, M., et al.: Parametric estimate of intensity inhomogeneities applied to MRI. TMI **19** (2000) 153–165

A Statistical Parts-Based Appearance Model of Inter-subject Variability

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

[1] Centre for Intelligent Machines, McGill University, Montréal, Canada
(mtoews, arbel)@cim.mcgill.ca
[2] McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montréal, Canada
louis@bic.mni.mcgill.ca

Abstract. In this article, we present a general statistical parts-based model for representing the appearance of an image set, applied to the problem of inter-subject MR brain image matching. In contrast with global image representations such as active appearance models, the parts-based model consists of a collection of localized image parts whose appearance, geometry and occurrence frequency are quantified statistically. The parts-based approach explicitly addresses the case where one-to-one correspondence does not exist between subjects due to anatomical differences, as parts are not expected to occur in all subjects. The model can be learned automatically, discovering structures that appear with statistical regularity in a large set of subject images, and can be robustly fit to new images, all in the presence of significant inter-subject variability. As parts are derived from generic scale-invariant features, the framework can be applied in a wide variety of image contexts, in order to study the commonality of anatomical parts or to group subjects according to the parts they share. Experimentation shows that a parts-based model can be learned from a large set of MR brain images, and used to determine parts that are common within the group of subjects. Preliminary results indicate that the model can be used to automatically identify distinctive features for inter-subject image registration despite large changes in appearance.

1 Introduction

Consider the task of inter-subject registration, i.e. determining correspondence between images of different subjects of a population. The task is one of great importance to the medical imaging community, as it lies at the heart of understanding how individuals vary within a population, but at the same time poses a significant challenge due to the problem of inter-subject variability. No two subjects are identical - structures may exhibit significant variation from one subject to the next or may simply not exist in all subjects. It is reasonable to expect that approaches attempting to determine a one-to-one mapping between

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 232–240, 2006.

different subjects may perform poorly in locations where such a mapping may not exist. Authors have attempted to identify points or regions where registration can be expected to perform well. Salient feature detection approaches are an option [8,9], but are generally ineffective due to inter-subject variability, as features extracted in one subject may not exist in the other. Features can be identified over a set of aligned subjects [5], but again one cannot know if these features will be relevant to new images.

We hypothesize that in order to determine the image regions likely to register well, one must first learn the degree of variability to be expected within the imaging context, and bring this knowledge to bear when registering new images. In this paper, we present a new probabilistic approach to automatically identifying image regions likely to result in meaningful inter-subject registration. Our approach is two-fold. First, off-line, we present a learning approach to building a parts-based appearance model of the image set in question (e.g. MR brain images). The model statistically quantifies the appearance, geometry and occurrence frequency of localized image parts which occur with statistical regularity over a set of training subjects. Second, we fit this model to new images to be registered, thereby identifying instances of learned model parts in the images. Statistically distinct image regions arising from the same model parts in both images are good candidates for registration.

The main contribution of this paper is a new parts-based statistical model of appearance that is based on localized image regions, as opposed to the global modes of image variation popular in the medical imaging literature [3,10]. The local nature of the method makes it possible to explicitly model situations where one-to-one correspondence does not exist between subjects due inter-subject variation, as model parts are not expected to (and typically do not) occur in all subjects. The model can be learned from large set of training images via a fully automatic machine learning algorithm and robustly fit to new images, all in the presence of significant inter-subject variation. The second contribution is a method of using this model to automatically determine the regions most likely to lead to meaningful registration between new images.

Our model is based on generic scale-invariant image features, and is general enough to be applied in a wide variety of image contexts. We present experiments based on T1-weighted MR brain images from the ICBM152 data set [2], which consists of 152 volumes of 88 male and 66 female normal subjects, aged 24.6 ± 4.8 years. Preliminary results indicate that learning is able to automatically identify a set of image parts, which can be used to determine commonality in new subjects exhibiting significant differences in appearance. The remainder of this paper is organized as follows: in Section 2 we present the parts-based model, in Section 3 we describe model learning and fitting, in Section 4 we present experimentation involving inter-subject registration, and in Section 5 we conclude with a discussion.

2 A Statistical Parts-Based Appearance Model

Identifying the regions most likely to match well between two different subjects requires first learning what image structure occurs with regularity in the particular imaging context, including the range of geometrical and appearance variability to be expected. To this end, we present a statistical model based on localized image *parts*, which here are image regions derived from generic, automatically detected scale-invariant features. A model based on such features is attractive because it can be automatically applied in a wide variety of contexts, as opposed to approaches based on special-purpose detectors for specific image structures (i.e. a particular sulcus) or manual landmark selection, which is tedious for large data sets and prone to inter-rater variation, as a human must decide which landmarks are optimal, how many landmarks must be used, etc.

Scale-invariant features [8,9] offer an improvement over simple features such as corners [6], as they localize salient image patterns in scale in addition to image translation. They can be extracted in a wide variety of image contexts, and are robust to variation in intensity, in addition to geometrical deformations such as translation, orientation, scale and affine transformations. Detection and correspondence of scale-invariance features are both fast: features can be extracted efficiently from image pyramids and matched after normalizing feature image content with respect to feature geometry, removing the need to perform an explicit search over deformation parameters. Feature geometrical parameters such as location x, orientation θ and scale σ recovered in the extraction process can be used formulate multiple independent hypotheses as to the geometrical transform between images, leading to occlusion and noise-resistant correspondence. In addition, scale-invariant features can be extracted from a variety of different image properties such as blobs [8], edges [9], phase [1] and entropy [7], and used in conjunction with each other.

Despite their attractive properties, scale-invariant features are rarely used for inter-subject registration because of the fact that, in general, the same features cannot be extracted in images of different subjects due to inter-subject variability. As a result, the number of features located is typically insufficient for reliably determining correspondence between different subjects. Although this fact can be considered a shortcoming of automatic feature detection, we argue in this paper that it merely reflects the difficulty of inter-subject registration, and indicates that one-to-one correspondence may not generally exist between all subjects. Recently, statistical models of features learned from a set of images have emerged in order to address these difficulties [4,12]. The parts-based model presented in this section is based on statistically quantifying the occurrence frequency, appearance and geometry of features over a large set of subjects, thereby learning a set of parts that can be reliably matched between images of different subjects.

2.1 Model Components

Our parts-based model consists of a set of parts m_i within a common reference frame o. Model parts are scale-invariant features denoted as $m_i : \{m_i^b, m_i^g, m_i^a\}$

representing the occurrence, geometry and appearance of a scale-invariant feature within an image. Feature occurrence m_i^b is a binary random variable representing the probability of feature presence (or absence) in an image. Feature geometry $m_i^g : \{x_i, \theta_i, \sigma_i\}$ is an oriented region in $\mathbb{R}^N$ image space, represented by N-parameter location x_i, an $N-1$ parameter orientation θ_i, and a scale σ_i. Feature appearance m_i^a describes the image content at region m_i^g, and can generally be parameterized in a number of ways, such as principle components.

Reference frame $o : \{o^b, o^g\}$ represents the occurrence and geometry of a common reference frame relating parts. o^b is a binary random variable indicating the presence/absence of the reference frame, who's significance will be made clear in the discussion of model learning and fitting. o^g is parameterized in the same manner as scale-invariant feature geometry m^g, and serves as a common geometrical frame within which different features m_i and m_j can be considered as statistically independent (i.e. knowing o^g, feature variation is localized to a neighborhood around an expected value of m_i). Within the context of MRI brain registration, a well-known definition of o^g is the midplane line defining the Talairach stereotactic space [11], which passes from the superior aspect of the anterior commissure to the inferior aspect of the posterior commissure. Figure 1 illustrates the relationship between o and m_i.

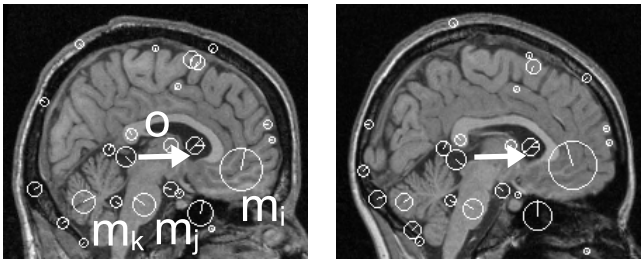


Fig. 1. Scale-invariant features and reference frames in sagittal slices of T1-weighted MR brain images of two subjects. Features m_i, illustrated as white circles inset by radial lines, are oriented regions consisting of a location x_i, orientation θ_i and scale σ_i. Reference frame o, illustrated as a white arrow, represents the projection of the Talairach AC-PC line onto the slice. Feature occurrence, appearance and geometry with respect to the reference frame can be quantified statistically via a parts-based model learned from a large set of subjects.

2.2 Probabilistic Model Formulation

Our model consists of a set of N model parts $\{m_i\}$, which when observed in a new image can be used infer the reference frame o. Assuming that parts m_i are conditionally independent given o, the posterior probability of o given $\{m_i\}$ can be expressed using Bayes rule as:

$$p(o|\{m_i\}) = \frac{p(o)p(\{m_i\}|o)}{p(\{m_i\})} = \frac{p(o)\prod_i^N p(m_i|o)}{p(\{m_i\})}, \tag{1}$$

where $p(o)$ is a prior over reference frame geometry and occurrence and $p(m_i|o)$ is the likelihood of feature m_i given o. Our model focuses principally on the likelihood term $p(m_i|o)$, which can be expressed as:

$$p(m_i|o) = p(m_i^a, m_i^b|o)p(m_i^g|o) = p(m_i^a|m_i^b)p(m_i^b|o^b)p(m_i^g|o^b, o^g), \quad (2)$$

under the assumptions that m^a and m^b are statistically independent of m^g given o, and that m^a and o are statistically independent given m^b.

Appearance likelihood $p(m_i^a|m_i^b)$ is represented as a multivariate Gaussian distribution in an appearance space and parameterized by mean and covariance μ_i^a, Σ_i^a. $p(m_i^b|o^b)$ is the probability of feature occurrence given reference frame occurrence, represented as a discrete multinomial distribution with event count parameters $\pi_i = \{\pi_i^1, \ldots, \pi_i^4\}$. Geometry likelihood $p(m_i^g|o^b, o^g)$ models the residual error of a linear transform from feature to reference frame geometry $m_i^g \rightarrow o^g$, and is represented as a Gaussian distribution with mean and covariance parameters μ_i^g, Σ_i^g. In order to characterize geometrical error in a scale-invariant manner, scale is transformed logarithmically, and translation is normalized by reference frame scale.

3 Model Learning and Fitting

Model learning involves estimating the parameters of the set of N features $\{m_i\}$ introduced in the previous section from a set of training images. Learning is based on a set of data vectors of the form $\{m_i^a, m_i^g, o^g\}$, where m_i^a and m_i^g are automatically extracted features and o^g is the labeled reference frame, here taken to be the AC-PC line of the Talairach stereotactic space [11]. Labeling o^g can be done by manually defining a line segment corresponding to o^g in images, or in an approximate manner via linear registration of MR volumes into the same stereotactic space, here we adopt the latter approach. Features are extracted and represented using the SIFT (scale-invariant feature transform) technique [8], based on an efficient implementation available online from the author, although a variety of other techniques could be used. Briefly, SIFT features are extracted as maxima/minima in a difference-of-Gaussian scale space pyramid, determining feature geometry m_i^g. The SIFT appearance representation m_i^a is a 128-value vector, corresponding to bins of a histogram of image first derivatives quantized into 8x4x4=128 bins over orientation and (x,y) position.

Prior to learning, feature geometries are normalized spatially wrt the reference frame $\{m_i^a, m_i^g, o^g\} \rightarrow \{m_i^a, \bar{m_i}^g\}$. Learning begins by clustering data vectors according to normalized geometry, thereby determining μ_i^g, Σ_i^g. Occurrence and appearance parameters π_i and μ_i^a, Σ_i^a are then estimated simultaneously such that the likelihood ratio $\frac{p(m_i^{b=1}|o^{b=1})}{p(m_i^{b=1}|o^{b=0})}$ is maximized. Note that this ratio represents a measure of the *distinctiveness* of a particular feature within the reference frame. As the number of clusters N is unknown *a priori*, clustering is achieved by growing variance estimates around individual vectors, instead of through algorithms such as K-means. After learning, features can be automatically ranked according

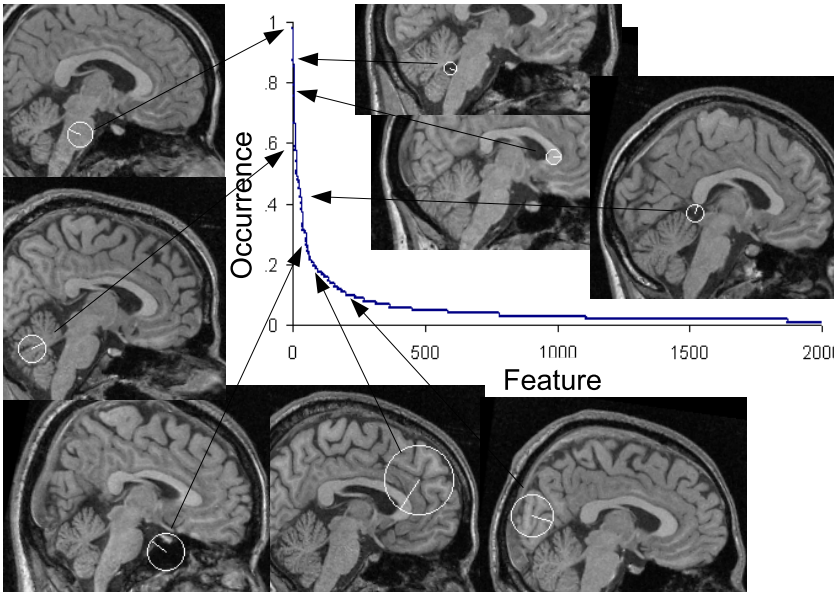


Fig. 2. A graph of features sorted by occurrence frequency $p(m_i^{b=1}|o^{b=1})$. The images illustrate features occurring at indicated frequency within a sample of 102 sagittal slices brains from the ICBM152 data set [2]. Note that feature occurrence drops off sharply, indicating that a relatively small number of features are common to all brains, whereas a large number of features are specific to a small number of brains.

to distinctiveness or occurence frequency, as illustrated in Figure 2. Having quantified part appearance, geometry and occurrence frequency, features observed in new images can be fit to the model and assessed as to their usefulness for tasks such as inter-subject registration. Highly distinctive features indicate good candidate regions for registration, whereas poorly distinctive features represent ambiguous or subject-specific characteristics.

Once the model has been learned, it can be fit to a new image by localizing the reference frame o. Unlike other registration/fitting techniques based on iterative algorithms which tend to go awry when started outside a 'capture radius' of the optimal solution, our model can be fit globally. In order to determine o in a new image, features extracted from the new image are matched to those in the model. Each pairing of an image feature to a model feature results in a hypothesis as to o^g in the new image. We are interested in evaluating whether these hypotheses are the result of a true model instance or random noise, i.e. $o = \{o^g, o^{b=1}\}$ or $\overline{o} = \{o^g, o^{b=0}\}$. These two possibilities can be compared via a Bayes decision ratio:

$$\gamma(o) = \frac{p(o|\{m_i\})}{p(\overline{o}|\{m_i\})} = \frac{p(o)}{p(\overline{o})} \prod_{i=1}^{N} \frac{p(m_i|o)}{p(m_i|\overline{o})}, \tag{3}$$

where high $\gamma(o)$ indicates the presence of a model, and $\frac{p(o)}{p(\overline{o})}$ is a constant expected ratio of true to false model instances. Fitting is performed by identifying o^g maximizing $\gamma(o)$, by considering clusters of model features that support similar hypotheses.

4 Application to Inter-subject Registration

Intuitively, fitting the model to a new image can be seen as describing the new image in terms of a unique combination of learned model parts. Images of different subjects can be compared in terms of the model parts that they share, the key notion being that any given pair of subjects shares a unique set of image parts. This information can potentially be used in a number of ways, for example to drive inter-subject registration in regions where images are known to have statistically similar content, or to cluster subject images that share similar image content. In this section, we wish explore how the features derived from the model can be applied to the task of inter-subject registration.

To apply the model to inter-subject image registration, we examine several cases where the model parts automatically learned through training are matched between new subjects not used in model learning. First, the model is fit individually to each new subject to be registered, identifying model features present in each subject. Then 10 model features of one subject are selected to be registered to the next, in two different ways: 1) selecting the most distinct model features in *one subject* similar to the strategy proposed by [5] and 2) selecting the most distinct model features common to *both subjects*. Registration is then performed by determining feature displacements from one image to the

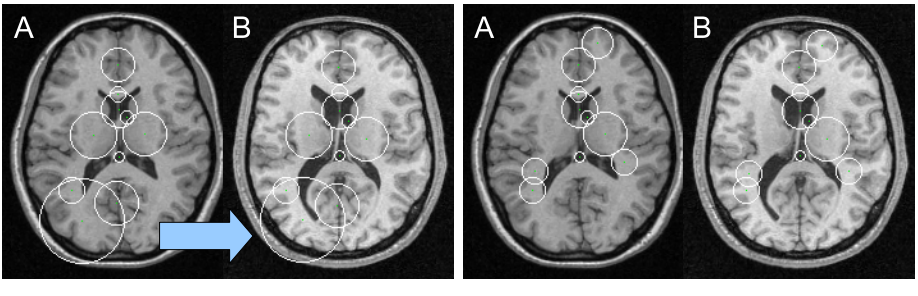


Fig. 3. Illustrating model-based feature selection for inter-subject registration. In each image pair, 10 features from subject A are selected in to be registered to subject B. On the left, the most distinct model features of subject A are used, on the right the most distinct features common to both subjects A and B are used. Notice that a valid registration solution may not exist in the lower left region, due to an enlarged ventricle of subject B. Basing registration on features common to both subjects, this ambiguous region can be avoided, resulting in more meaningful registration in the presence of inter-subject variability.

next, based on a sum-of-squared-difference similarity measure regularized by an elastic prior between features. When registering normal subjects, both feature selection techniques result in reasonable registration, as most subjects share similar features. In the presence of significant inter-subject variability however, selecting features common to both subjects allows registration to avoid regions in which a valid solution may not exist, such as the enlarged ventricle illustrated in Figure 3.

5 Discussion

In this paper, we presented a statistical parts-based appearance model, where a set of images is represented by a collection of automatically-extracted generic parts. The parts-based approach is unique in that it models variation on a local scale, and as such is able to explicitly model situations where one-to-one correspondence does not exist between subjects. Experimentation showed that model parts can be identified and quantified statistically by a fully automatic learning algorithm, and robustly fit to new images, all in the presence of significant inter-subject variability. In addition, preliminary results demonstrated that the model can be used to identify statistically distinct parts common to different subjects, indicating image structure that can be matched with higher reliability for the task of inter-subject registration. In general, parts-based modeling represents a powerful tool for understanding the variation in local anatomical structure over a large set subjects, and offers many avenues for future exploration. Part statistics could potentially serve as indicators of abnormal conditions such as pathology. A parts-based framework could potentially be used to relate anatomical structures across different species. These encouraging results and future directions should naturally lead to the development of robust scale-invariant feature detectors in 3-D and 4-D, and the application of the model in a wide variety of contexts.

References

1. G. Carneiro and A.D. Jepson, *Multi-scale phase-based local features*, CVPR, vol. 1, 2003, pp. 736–743.
2. D. L. Collins, N.J Kabani, and A.C. Evans, *Automatic volume estimation of gross cerebral structures*, 4th International Conference of Functional Mapping of the Human Brain (A. Evans, ed.), 1998.
3. T.F. Cootes, G.J. Edwards, and C.J. Taylor, *Active appearance models*, Pattern Analysis and Machine Intelligence **23** (2001), no. 6, 681–684.
4. R. Fergus, P. Perona, and A. Zisserman, *Object class recognition by unsupervised scale-invariant learning*, CVPR, 2003, pp. 264–271.
5. Wu Guorong, Feihu Qi, and Dinggang Shen, *Learning best features for deformable registration of mr brains*, MICCAI, 2005.
6. C. Harris and M. Stephens, *A combined corner and edge detector*, Proceedings of the 4th Alvey Vision Conference, 1988, pp. 147–151.
7. T. Kadir and M. Brady, *Saliency, scale and image description*, IJCV **45** (2001), no. 2, 83–105.

8. David G. Lowe, *Distinctive image features from scale-invariant keypoints*, IJCV **60** (2004), no. 2, 91–110.
9. K. Mikolajczyk and C. Schmid, *Scale and affine invariant interest point detectors*, IJCV **60** (2004), no. 1, 63–86.
10. Daniel Rueckert, Alejandro F. Frangi, and Julia A. Schnabel, *Automatic construction of 3-d statistical deformation models of the brain using nonrigid registration*, TMI **22** (2003), no. 8, 1014–1025.
11. J. Talairach and P. Tournoux, *Co-planar stereotactic atlas of the human brain: 3-dimensional proportional system: an approach to cerebral imaging*, Georg Thieme Verlag, Stuttgart, 1988.
12. M. Toews and T. Arbel, *Detection over viewpoint via the object class invariant*, ICPR, 2006.

The Entire Regularization Path for the Support Vector Domain Description

Karl Sjöstrand[1,2] and Rasmus Larsen[1]

[1] Informatics and Mathematical Modelling, Technical University of Denmark
[2] Department of Radiology, VAMC, University of California-San Francisco, USA
kas@imm.dtu.dk, rl@imm.dtu.dk

Abstract. The support vector domain description is a one-class classification method that estimates the shape and extent of the distribution of a data set. This separates the data into outliers, outside the decision boundary, and inliers on the inside. The method bears close resemblance to the two-class support vector machine classifier. Recently, it was shown that the regularization path of the support vector machine is piecewise linear, and that the entire path can be computed efficiently. This paper shows that this property carries over to the support vector domain description. Using our results the solution to the one-class classification can be obtained for any amount of regularization with roughly the same computational complexity required to solve for a particularly value of the regularization parameter. The possibility of evaluating the results for any amount of regularization not only offers more accurate and reliable models, but also makes way for new applications. We illustrate the potential of the method by determining the order of inclusion in the model for a set of corpora callosa outlines.

1 Introduction

The support vector domain description (SVDD) [1,2] is a one-class classification method for unlabeled data that is closely related to the support vector machine (SVM) [3] for labeled (training) data with two or more classes. While SVM separates two classes using a hyperplane, SVDD separates data from outliers using a hypersphere. The idea is to find the minimal sphere that encapsulates the data, allowing for some points to be outside the boundary. Formally, this amounts to,

$$\min_{R^2,\mathbf{a},\xi_i} \sum_i \xi_i + \lambda R^2, \quad \text{subject to} \quad (\mathbf{x}_i - \mathbf{a})(\mathbf{x}_i - \mathbf{a})^T \leq R^2 + \xi_i, \quad \xi_i \geq 0 \quad \forall i, \quad (1)$$

where $\mathbf{X} = [\mathbf{x}_1^T \dots \mathbf{x}_n^T]^T$ is the $n \times p$ data matrix with $\mathbf{x}_i \in \mathbb{R}^p$, R is the radius, $\mathbf{a}$ is the center of the sphere, and ξ_i is the amount by which point i is allowed to be outside the sphere. The parameter λ determines the amount of regularization. A large value of λ puts focus on the radius, leading to a smaller sphere while the ξ_i are allowed to grow large. If λ is small, the resulting sphere will be large

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 241–248, 2006.

while the total distance of outlying points shrinks. The setup in Equation 1 is convex and can be solved using Lagrange multipliers $\alpha_i \geq 0$ and $\gamma_i \geq 0$,

$$L_p : \sum_i \xi_i + \lambda R^2 + \sum_i \alpha_i (\mathbf{x}_i \mathbf{x}_i^T - 2\mathbf{a}\mathbf{x}_i^T + \mathbf{a}\mathbf{a}^T - R^2 - \xi_i) - \sum_i \gamma_i \xi_i. \quad (2)$$

At the minimum, the derivative of each variable is zero, giving

$$\frac{\partial L_p}{\partial R^2} = 0 \Leftrightarrow \lambda = \sum_i \alpha_i, \quad (3)$$

$$\frac{\partial L_p}{\partial \mathbf{a}} = 0 \Leftrightarrow \mathbf{a} = \frac{\sum_i \alpha_i \mathbf{x}_i}{\sum_i \alpha_i} = \frac{\sum_i \alpha_i \mathbf{x}_i}{\lambda}, \quad (4)$$

$$\frac{\partial L_p}{\partial \xi_i} = 0 \Leftrightarrow \alpha_i = 1 - \gamma_i. \quad (5)$$

Equation 5 and the fact that $\alpha_i \geq 0, \gamma_i \geq 0, \forall i$, gives that $0 \leq \alpha_i \leq 1$. Furthermore, we have the Karush-Kuhn-Tucker complimentary conditions,

$$\alpha_i (\mathbf{x}_i \mathbf{x}_i^T - 2\mathbf{a}\mathbf{x}_i^T + \mathbf{a}\mathbf{a}^T - R^2 - \xi_i) = 0, \quad (6)$$

$$\gamma_i \xi_i = 0. \quad (7)$$

Equations 5, 6 and 7 give that $\alpha_i = 1$ for points outside the sphere and $\alpha_i = 0$ for points on the inside. By continuity, α_i will travel from 1 to 0 as point i passes the boundary from outside the sphere to the inside. Inserting Equations (3-5) into (2) gives the Wolfe dual form which is to be maximized w.r.t. (3-5),

$$L_w : \max_{\alpha_i} \sum_i \alpha_i \mathbf{x}_i \mathbf{x}_i^T - \frac{1}{\lambda} \sum_i \sum_j \alpha_i \alpha_j \mathbf{x}_i \mathbf{x}_j^T \quad : \quad 0 \leq \alpha \leq 1, \quad \sum_i \alpha_i = \lambda. \quad (8)$$

This is a quadratic optimization problem with linear constraints. As such, it can be solved using some quadratic programming algorithm. This is a slight reformulation of the original setup [1] which uses a regularization parameter $C = 1/\lambda$. As in [4] we favor the description above since $0 \leq \alpha \leq 1$ instead of $0 \leq \alpha \leq C$ which facilitates the interpretation of the coefficient paths $\alpha_i(\lambda)$.

Equation 3 determines the valid range for the regularization parameter to $\lambda \in [0, n]$.

For most data sets, a hypersphere is an unsuitable representation of the data. Increasing the dimensionality using basis expansions $h(\mathbf{x})$ allows for more flexible decision boundaries. Replacing $\mathbf{x}$ by $h(\mathbf{x})$ in Equation 8, we note that the dual can be expressed in terms of inner products $\langle h(\mathbf{x}_i), h(\mathbf{x}_j) \rangle$. These can be replaced by $K_{i,j} = K(\mathbf{x}_i, \mathbf{x}_j)$, where K is some suitable kernel function,

$$L_w : \max_{\alpha_i} \sum_i \alpha_i K_{i,i} - \frac{1}{\lambda} \sum_i \sum_j \alpha_i \alpha_j K_{i,j} \quad : \quad 0 \leq \alpha \leq 1, \quad \sum_i \alpha_i = \lambda. \quad (9)$$

In the remainder of this paper, the more general kernel notation will be used.

The squared distance from the center of the sphere to a point $\mathbf{x}$ is,

$$f(\mathbf{x}) = \|h(\mathbf{x}) - \mathbf{a}\|^2 = K(\mathbf{x},\mathbf{x}) - \frac{2}{\lambda}\sum_i \alpha_i K(\mathbf{x},\mathbf{x}_i) + \frac{1}{\lambda^2}\sum_i\sum_j \alpha_i\alpha_j K_{i,j} \quad (10)$$

The squared radius of the sphere can therefore be written $R^2 = f(\mathbf{x}_k)$, where index k belongs to any point on the boundary ($\alpha_k \in (0,1)$). Note that $f(\mathbf{x})$, R^2, α_i and index k are all dependent on λ.

2 Computing the Regularization Path

In this section we will prove that the coefficient path of each α_i is piecewise linear in λ, and purpose an algorithm for their calculation using standard matrix algebra. The increased interest in coefficient paths and their computation originates from the seminal work by Efron et al. [5], where the regularization path of the LASSO regression algorithm is derived, leading to a highly efficient algorithm. These results have since been generalized to hold for a range of regularized problems [6,7]. Specifically, it holds for support vector machines as described by Hastie et al. [4]. Due to the alikeness of SVMs and the SVDD, the following derivation of the regularization path for the SVDD is noticeably similar to that of SVMs.

Define by $\mathcal{I}$, $\mathcal{O}$ and $\mathcal{B}$ the sets containing indices i corresponding to points on the inside, outside and on the boundary respectively, and let $n_\mathcal{I}$, $n_\mathcal{O}$, and $n_\mathcal{B}$ be the number of elements in these sets. The set $\mathcal{A} = \mathcal{I} \cup \mathcal{O} \cup \mathcal{B}$ contains the indices of all points.

As discussed above, $\alpha_i = 1$ for $i \in \mathcal{O}$, $\alpha_i = 0$ for $i \in \mathcal{I}$, and $0 < \alpha_i < 1$ for $i \in \mathcal{B}$. There are four types of events where these sets change.

1. Point i leaves $\mathcal{B}$ and joins $\mathcal{I}$; $\alpha_i \in (0,1) \longrightarrow \alpha_i = 0$.
2. Point i leaves $\mathcal{B}$ and joins $\mathcal{O}$; $\alpha_i \in (0,1) \longrightarrow \alpha_i = 1$.
3. Point i leaves $\mathcal{O}$ and joins $\mathcal{B}$; $\alpha_i = 1 \longrightarrow \alpha_i \in (0,1)$.
4. Point i leaves $\mathcal{I}$ and joins $\mathcal{B}$; $\alpha_i = 0 \longrightarrow \alpha_i \in (0,1)$.

To determine which set a point $\mathbf{x}$ belongs to, we define a decision function,

$$g(\mathbf{x}) = f(\mathbf{x}) - f(\mathbf{x}_k) = K(\mathbf{x},\mathbf{x}) - K_{k,k} - \frac{2}{\lambda}\sum_i \alpha_i\big(K(\mathbf{x},\mathbf{x}_i) - K_{k,i}\big),\ k \in \mathcal{B}, \quad (11)$$

which has $g = 0$ for $\mathbf{x}$ on the boundary, $g < 0$ for $\mathbf{x}$ inside and vice versa.

The algorithm starts at $\lambda = n$, corresponding to the minimal sphere radius with $\mathcal{O} = \mathcal{A}$ and $\alpha_i = 1$, $\forall i$.

From an arbitrary configuration of $\mathcal{I}$, $\mathcal{O}$ and $\mathcal{B}$, λ is allowed to decrease until the next event occurs. As in [4], let λ_l be the value of the regularization parameter at step l. While $\lambda_l < \lambda < \lambda_{l+1}$, the sets remain static. Hence, $g(\mathbf{x}_m) = 0, \forall\, m \in \mathcal{B}$ in this interval. Using this, Equation 11 can be expanded and rearranged into

$$\sum_{i\in\mathcal{B}} \alpha_i(K_{m,i} - K_{k,i}) = \frac{\lambda}{2}(K_{m,m} - K_{k,k}) - \sum_{i\in\mathcal{O}}(K_{m,i} - K_{k,i})\ \ \forall m \in \mathcal{B}. \quad (12)$$

This results in $n_{\mathcal{B}}$ equations with $n_{\mathcal{B}}$ unknowns $\alpha_i, i \in \mathcal{B}$. However, for $m = k$, it is seen that (12) degenerates into $0 = 0$, making the system of equations rank deficient. We therefore replace the equation for $m = k$ by the auxiliary condition in Equation 3.

This procedure can be summarized in matrix form. Let $\mathbf{Y}$ be an $n \times n$ matrix where $\mathbf{Y}_{i,j} = K_{i,j} - K_{k,j},\ \forall\,(i,j) \in \mathcal{A}$ and let $\mathbf{y}$ be a length n vector with $y_i = K_{i,i} - K_{k,k}\ \forall i \in \mathcal{A}$. With the obvious definitions of submatrices, Equation 12 can be written

$$\mathbf{Y}_{\mathcal{B},\mathcal{B}}\alpha_{\mathcal{B}} = \frac{\lambda}{2}\mathbf{y}_{\mathcal{B}} - \mathbf{Y}_{\mathcal{B},\mathcal{O}}\mathbf{1}_{n_{\mathcal{O}}}, \tag{13}$$

where $\mathbf{1}_{n_{\mathcal{O}}}$ is a vector of ones of length $n_{\mathcal{O}}$. This expression can be expanded to include the conditions $\alpha_i = 0$ for $i \in \mathcal{I}$ and $\alpha_i = 1$ for $i \in \mathcal{O}$. It also needs to be augmented to replace the degenerate equation corresponding to index k with the relation from Equation 3. We will now define matrices that implement this.

Let $\mathcal{B}_{-k}$ be the boundary set with index k removed. Let $\mathbf{Z}$ be the $n \times n$ identity matrix with $\mathbf{Z}_{\mathcal{B}_{-k},\mathcal{B}_{-k}} = \mathbf{Y}_{\mathcal{B}_{-k},\mathcal{B}_{-k}}$ and $\mathbf{Z}_{k,\mathcal{A}} = \mathbf{1}_n^T$. Let $\mathbf{z}$ be the length n zero vector with $\mathbf{z}_{\mathcal{B}_{-k}} = \mathbf{y}_{\mathcal{B}_{-k}}$ and $z_k = 2$. Finally, let $\mathbf{W}$ be the $n \times n$ zero matrix with $\mathbf{W}_{\mathcal{B}_{-k},\mathcal{O}} = -\mathbf{Y}_{\mathcal{B}_{-k},\mathcal{O}}$ and $\mathbf{W}_{\mathcal{O},\mathcal{O}} = \mathbf{I}_{n_{\mathcal{O}}}$ where $\mathbf{I}_{n_{\mathcal{O}}}$ is the identity matrix of size $n_{\mathcal{O}}$. The complete system of n equations for n unknowns is then

$$\mathbf{Z}\alpha = \frac{\lambda}{2}\mathbf{z} + \mathbf{W}\mathbf{1}_n. \tag{14}$$

Providing $\mathbf{Z}$ is invertible, the resulting expression for α becomes,

$$\alpha = \frac{\lambda}{2}\mathbf{Z}^{-1}\mathbf{z} + \mathbf{Z}^{-1}\mathbf{W}\mathbf{1}_n \equiv \lambda\mathbf{p} + \mathbf{q}, \tag{15}$$

an expression that is linear in λ.

Now that the functional form of each coefficient between two events has been established, it remains to disclose the valid range $[\lambda_l, \lambda_{l+1}]$ of λ. That is, we wish to find λ_{l+1} at which the next event occurs. We treat each of the four types of events defined above separately.

The first event type occurs for α_i, $i \in \mathcal{B}$ when $\alpha_i \to 0$. By setting (15) equal to 0 and solving for each value of λ, we get,

$$\lambda_i = -\frac{q_i}{p_i}, \quad i \in \mathcal{B}. \tag{16}$$

Similarly for event two, $\alpha_i = 1$ when

$$\lambda_i = \frac{1 - q_i}{p_i}, \quad i \in \mathcal{B}. \tag{17}$$

For either of the other two event types to occur, a point i in either $\mathcal{I}$ or $\mathcal{O}$ must join the boundary. At this stage, $g(\mathbf{x}_i) = 0$. Inserting (15) into (11), the value of the decision function at point i for some value of λ can be expressed as

$$g(\mathbf{x}_i, \lambda) = y_i - 2\mathbf{Y}_{i,\mathcal{A}}(\mathbf{p} + \frac{1}{\lambda}\mathbf{q}). \tag{18}$$

To find the values of λ at which each point joins the boundary, $g(\mathbf{x}_i, \lambda_i) = 0$ is solved for λ_i,

$$\lambda_i = \frac{2\mathbf{Y}_{i,\mathcal{A}}\mathbf{q}}{y_i - 2\mathbf{Y}_{i,\mathcal{A}}\mathbf{p}}. \tag{19}$$

Out of the candidates $\{\lambda_i\}$ for λ_{l+1}, the largest candidate smaller than λ_l must be the point at which the sets first change. Therefore, $\lambda_{l+1} = \max_i \lambda_i$ subject to $\lambda_i < \lambda_l$.

There is one final consideration. Event 1 may at any stage of the algorithm lead to the boundary set $\mathcal{B}$ becoming empty, resulting in a violation of Equation 3. One or more points from $\mathcal{O}$ must therefore join $\mathcal{B}$ concurrently. The calculation of candidates for λ_{l+1} in (19) will fail in this case, as a consequence of the new point not being placed on the current boundary. This behavior forces a discontinuity in the radius function, which must increase discretely to encompass the next point. Since $\alpha(\lambda)$ is a continuous function, Equation 4 shows that the position of the hypersphere center $\mathbf{a}(\lambda)$ is also continuous. Hence, despite the discontinuity of the boundary function, the next point to join $\mathcal{B}$ can be established by finding the point in $\mathcal{O}$ with the smallest radius,

$$\min_{i\in\mathcal{O}} f(\mathbf{x}_i) = \min_{i\in\mathcal{O}} K_{i,i} - \frac{2}{\lambda}K_{i,\mathcal{A}}\alpha + \frac{1}{\lambda^2}\alpha^T K_{\mathcal{A},\mathcal{A}}\alpha. \tag{20}$$

The entire process is summarized in Algorithm 1.

Algorithm 1. SVDD coefficient paths

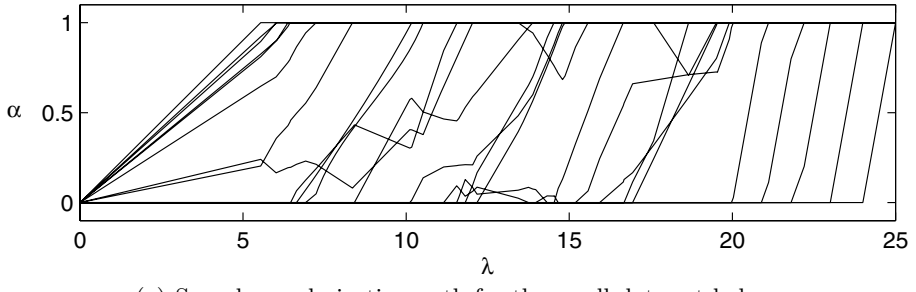
```
Initialize λ = n and α_i = 1 ∀i.
while λ > 0 do
    if n_B = 0 then
        Add index i to boundary set B that satisfies (20).
        Remove i from O.
    end if
    Given sets I, O and B, compute p = Z^{-1}z/2 and q = Z^{-1}W1_n.
    Calculate λ candidates according to event 1 using (16).
    Calculate λ candidates according to event 2 using (17).
    Calculate λ candidates according to event 3 using (19) with i ∈ O.
    Calculate λ candidates according to event 4 using (19) with i ∈ I.
    Choose candidate λ_{l+1} with the largest value smaller than λ_l.
    Calculate new coefficients, α = λ_{l+1}p + q.
    Update sets accordingly.
end while
```

3 Examples

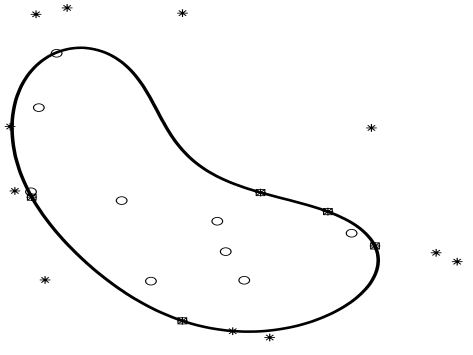
The algorithm was implemented using MATLAB and tested on both synthetic and real data sets. For the results presented here, a Gaussian kernel with $K_{i,j} = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/\gamma)$ has been used. High values of γ produce smooth and coherent

decision boundaries while smaller values give more wiggly and clustered results. The choice of this parameter is application specific and is not discussed in this paper. For the examples shown here, $\gamma = 1$ is used which results in smooth, but not excessively constrained boundaries.

Figure 1 shows a small data set of 25 points in two dimensions. Figure 1(a) shows the resulting regularization path for λ values ranging from 0 to $n = 25$. The piecewise linearity of the paths is apparent. Interpreting the paths for growing λ, it can be seen that the most common behavior for a point is to join the boundary from the inside and shortly after leave for the outside. However, it can be seen that there are exceptions to this rule, despite the constant shrinkage of the hypersphere. This is an effect of the movement of the sphere center.



(a) Sample regularization path for the small data set below.



(b) Decision boundary at $\lambda = 13$. Circles, stars and squares respectively denote inliers, outliers and points on the boundary.

Fig. 1. Example description of 25 points in two dimensions with the corresponding regularization path

The second experiment is an application that uses the asset of knowing the entire regularization path. The goal is to order a large set of Procrustes aligned shapes in ascending order according to the density of the corresponding distribution at each observation. This is done using two approaches. The first is based on

successive maximization of Mahalanobis distance. At each step, the observation with the largest distance w.r.t. the current data set is removed. For n shapes, this is performed $n-1$ times, thus establishing an ordering. The second method uses the SVDD and its regularization path. The order is established directly from $\mathcal{O}$ as λ grows from 0 to n. The data consists of 582 outlines of the mid-sagittal cross-section of the corpus callosum brain structure. This data set is part of the LADIS (Leukoaraiosis and DISability) study [8], a pan-European study involving 12 hospitals and more than 600 patients. Figures 2 and 3 show the first and last twelve observations of each ordering. The Mahalanobis distance measure is based on the shape of the covariance matrix and assumes an ellipsoidal distribution. Due to the use of kernels, the SVDD is able to model more complex distributions, giving better estimates of the density at each observation. This is particularly apparent among the inliers in Figure 3. The variance is clearly lower for the SVDD-based ordering than for the Mahalanobis-based counterpart.

(a) First 12 selected (outliers)

(b) Last 12 selected (inliers)

Fig. 2. Ordering established by successive maximization of Mahalanobis distance

(a) First 12 selected (outliers)

(b) Last 12 selected (inliers)

Fig. 3. Ordering established by the SVDD regularization path. Note the increased dissimilarity among the outliers, as well as the increased similarity among later samples.

4 Discussion

The computational complexity of computing the entire path is low. Most of the effort goes into solving the linear system in (15). To increase efficiency, we solve (15) for points on the boundary only, i.e. using submatrices $\mathbf{Z}_{\mathcal{B},\mathcal{B}}$, $\mathbf{z}_{\mathcal{B}}$ and $\mathbf{W}_{\mathcal{B},\mathcal{O}}$.

The remaining values of α_i ($i \in \mathcal{I} \cup \mathcal{O}$) remain static. The resulting burden for finding $\mathbf{p}$ and $\mathbf{q}$ is roughly $O(n_{\mathcal{B}}^3)$ where typically $n_{\mathcal{B}} \ll n$. The most prominent addition to this is the work involved in finding λ_{l+1} which includes the multiplication of several length n vectors. In our experience, the number of iterations is generally less than $2n$, although more than $5n$ iterations is possible for very dense data sets. In comparison, algorithms for solving quadratic programming problems have $O(n^k)$, with k dependent on the choice of implementation.

Due to the exclusive use of kernels, the method handles data with many variables well. The memory usage level is mainly due to the matrix $\mathbf{Y}_{\mathcal{B},\mathcal{B}}$, which can grow large for data sets with many observations and the use of very unconstrained decision boundaries.

Knowledge of the entire regularization path is an important basis for picking the appropriate amount of regularization. As shown in this paper, information from the path itself can also be used directly as a data description method. Furthermore, the SVDD is the basis for other methods such as support vector clustering (SVC) [9], which may benefit from these results.

Acknowledgement

The authors extend their gratitude to the LADIS work group for supplying the corpus callosum data. In particular we acknowledge the annotation effort of Charlotte Ryberg and Egill Rostrup from the Danish Research Center for Magnetic Resonance, Copenhagen University Hospital, Hvidovre, Denmark.

References

1. Tax, D.M., Duin, R.P.: Support vector domain description. Pattern Recognition Letters **20**(11-13) (1999) 1191–1199
2. Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., Williamson, R.C.: Estimating the support of a high-dimensional distribution. Neural Computation **13** (2001) 1443–1471
3. Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer, New York (1995)
4. Hastie, T., Rosset, S., Tibshirani, R., Zhu, J.: The entire regularization path for the support vector machine. JMLR **5** (2004) 1391–1415
5. Efron, B., Hastie, T., Johnstone, I., Tibshirani, R.: Least angle regression. Annals of Statistics **32**(2) (2004) 407–451
6. Rosset, S., Zhu, J.: Piecewise linear regularized solution paths. Technical report, Stanford University (2003)
7. Rosset, S.: Tracking curved regularized optimization solution paths. NIPS (2004)
8. Pantoni, L., Basile, A.M., Pracucci, G., Asplund, K., Bogousslavsky, J., Chabriat, H., Erkinjuntti, T., Fazekas, F., Ferro, J.M., Hennerici, M., O'brien, J., Scheltens, P., Visser, M.C., Wahlund, L.O., Waldemar, G., Wallin, A., Inzitari, D.: Impact of age-related cerebral white matter changes on the transition to disability - the LADIS study: Rationale, design and methodology. Neuroepidemiology **24**(1-2) (2005) 51–62
9. Ben-Hur, A., Horn, D., Siegelmann, H.T., Vapnik, V.: Support vector clustering. Journal of Machine Learning Research **2** (2001) 125–137

A New Closed-Form Information Metric for Shape Analysis

Adrian Peter[1] and Anand Rangarajan[2]

[1]Dept. of ECE, [2]Dept. of CISE, University of Florida, Gainesville, FL

Abstract. Shape matching plays a prominent role in the analysis of medical and biological structures. Recently, a unifying framework was introduced for shape matching that uses mixture-models to couple both the shape representation and deformation. Essentially, shape distances were defined as geodesics induced by the Fisher-Rao metric on the manifold of mixture-model represented shapes. A fundamental drawback of the Fisher-Rao metric is that it is NOT available in closed-form for the mixture model. Consequently, shape comparisons are computationally very expensive. Here, we propose a new Riemannian metric based on generalized ϕ- entropy measures. In sharp contrast to the Fisher-Rao metric, our new metric is available in closed-form. Geodesic computations using the new metric are considerably more efficient. Discriminative capabilities of this new metric are studied by pairwise matching of corpus callosum shapes. Comparisons are conducted with the Fisher-Rao metric and the thin-plate spline bending energy.

1 Introduction

Shape analysis is a key ingredient to many medical imaging applications that seek to study the intimate relationship between the form and function of medical and biological structures. In particular, landmark-based deformable models have been widely used in quantified studies requiring size and shape similarity comparisons. A brief, cross-cutting survey of existing work in shape analysis illustrates several taxonomies and summaries. Shape deformation parameterizations range from Procrustean metrics [1] to spline-based models [2, 3], and from PCA-based modes of deformation [4] to landmark diffeomorphisms [5, 6]. Shape representations range from unstructured point-sets [7, 8] to weighted graphs [9] and include curves, surfaces and other geometric models. These advances have been instrumental in solidifying the shape analysis landscape. However, one commonality in virtually all of this previous work is the use of separate models for shape representation and deformation. In recent work, it was shown that shape representation and deformation can be unified within a probabilistic framework [10]. By using a mixture of densities to represent shapes, the authors in [10] were able to directly use the Fisher information matrix to establish an intrinsic, Riemannian metric and subsequently a shape distance measure—thus unifying both shape representation and deformation.

But are we always handcuffed to the Fisher-Rao Riemannian metric when trying to establish distances between parametric, probabilistic models? (Remember in this context the parametric models are used to represent shapes.) The metric's close connections to Shannon entropy and the concomitant use of Fisher information in parameter

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 249–256, 2006.

estimation have cemented it as the incumbent information measure. It has also been proliferated by research efforts in information geometry, where one can show its proportionality to popular divergence measures such as Kullback-Leibler. However, the algebraic form of the Fisher-Rao metric tensor makes it very difficult to use when applied to multi-parameter spaces like mixture models. For instance, it is not possible to derive closed-form solutions for the metric tensor or its derivative. To address many of these computational inefficiencies that arise when using the standard information metric, the current work introduces a new Riemannian metric based on the generalized notion of a ϕ-entropy functional. We take on the challenge of improving the initial model as presented in [10] by incorporating the notion of generalized information metrics as first shown by Burbea and Rao [11]. Section 2 provides a synopsis of their work on obtaining differential metrics using the ϕ-entropy functional in parametric probability spaces. In Section 3, we discuss the use of a specific ϕ-function that leads to the α-order entropy first introduced by Havrda and Charvát [12]. We show how this metric leads to closed-form solutions for the Christoffel symbols when using a Gaussian mixture models (GMM) for coupling shape representation and deformation, enabling almost an order of magnitude performance increase over the Fisher-Rao based solution. Section 4 applies the new metric to compute shape distances between corpus callosum data and provides comparative analysis with the Fisher-Rao metrics and the thin-plate spline bending energy. Some closed-form solutions to the new metric tensor are provided in the Appendix.

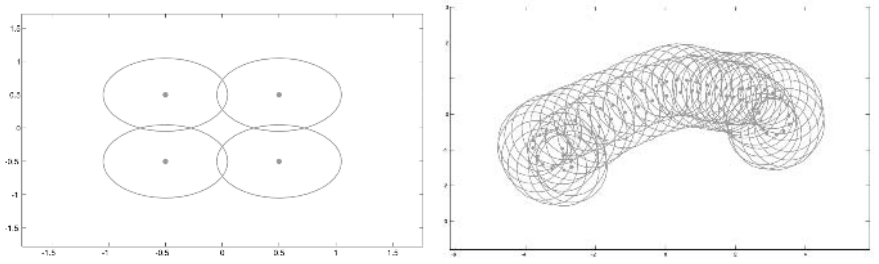
2 Differential Metrics for Parametric Distributions

The parametric, GMM representation for 2-D shapes is given by

$$p(\mathbf{x}|\theta) = \frac{1}{2\pi\sigma^2 K}\sum_{a=1}^{K}\exp\{-\frac{\|\mathbf{x}-\phi_a\|^2}{2\sigma^2}\} \tag{1}$$

where θ is our set of landmarks, $\phi_a = [\theta^{(2a-1)}, \theta^{(2a)}]^T$, $\mathbf{x} = [x^{(1)}, x^{(2)}]^T \in \mathbb{R}^2$ and equal weight priors are assigned to all components, i.e $\frac{1}{K}$ (Extensions to 3D are mathematically straightforward.) Basically, a shape with K landmarks is represented as a K-component GMM where the landmarks are the means of each component. The model $p(\mathbf{x}|\theta)$ lives on a statistical manifold M which has its local coordinates defined by the model parameters. Following [10], we choose to only place the Gaussian means on the manifold and assume an isotropic variance σ^2 for the landmarks (see examples in Figure 1).

Given that our representation model depends on parametric mixture models, it is important to understand how to obtain a metric on the parameter space. Rao established [13] that the Fisher information matrix satisfies the properties of a Riemannian metric on the statistical manifold M. This seminal work and the information matrix's relationship to the Shannon entropy have entrenched it as the metric tensor of choice when trying to establish a distance metric between two parametric models. However, Burbea and Rao went on to show that the notion of distances between parametric models can be extended to a large class of generalized metrics [11]. They defined the generalized *ϕ-entropy functional*



(a) (b)

Fig. 1. Examples of representation model. (a) The square shape is represented using four landmarks ($K = 4$) given by $\theta = \{[\phi_1 = (-0.5, 0.5)], [\phi_2 = (-0.5, 0.5)], [\phi_3 = (0.5, -0.5)], [\phi_4 = (0.5, 0.5)]\}$ and with $\sigma^2 = 0.15$. (b) Representation of corpus callosum with $K = 63$ and $\sigma^2 = 0.5$.

$$H_\phi(p) = -\int_\chi \phi(p) d\mathbf{x} \tag{2}$$

where χ is the measurable space (for our purposes $\mathbb{R}^2$), and ϕ is a C^2-convex function defined on $\mathbb{R}_+ \equiv [0, \infty)$. (For readability we will regularly replace $p(\mathbf{x}|\theta)$ with p.) The metric on the parameter space is obtained by finding the Hessian of (2) along a direction in its tangent space. Assuming sufficient regularity properties on $\theta = \{\theta^1, \ldots, \theta^n\}$, the direction can be obtained by taking the total differential of $p(\mathbf{x}|\theta)$ w.r.t θ

$$dp(\theta) = \sum_{k=1}^{n} \frac{\partial p}{\partial \theta^k} d\theta^k \ . \tag{3}$$

Hence the Hessian is defined as

$$\Delta_\theta H_\phi(p) = -\int_\chi \phi''(p)[dp(\theta)]^2 d\mathbf{x} \ . \tag{4}$$

This directly leads to the following differential metric satisfying Riemannian metric properties

$$ds^2_\phi(\theta) = -\Delta_\theta H_\phi(p) = \sum_{i,j=1}^{n} g^\phi_{i,j} d\theta^i d\theta^j \ , \tag{5}$$

where

$$g^\phi_{i,j} = \int_\chi \phi''(p)(\frac{\partial p}{\partial \theta^i})(\frac{\partial p}{\partial \theta^j}) d\mathbf{x} \ . \tag{6}$$

The metric tensor in (6) is called the *ϕ-entropy matrix.* We can define distances between probability densities by finding a geodesic between their parameter values as determined by (5). If one lets

$$\phi(p) = p \log p, \tag{7}$$

equation (2) becomes the familiar Shannon entropy and (6) yields the Fisher information matrix.

This section illustrated the tight coupling between entropy selection and its corresponding differential metric. In [10], the authors chose (7) and worked with resulting Fisher information matrix to compute geodesics between parametric GMM models, $p(\mathbf{x}|\theta_1)$ and $p(\mathbf{x}|\theta_2)$, which represented two shapes. However, computation of geodesics using this metric is very inefficient as they require numerical calculation of the integral in (6). We now discuss an alternative choice of ϕ that directly leads to a new Riemannian metric and enables us to derive closed-form solutions for (6).

3 α-Order Entropy Metric for Shape Analysis

Our desire to find a computationally efficient information metric was motivated by noticing that if the integral of the metric could be reduced to just a correlation between the partials of the density w.r.t θ^i and θ^j, i.e. $\int \frac{\partial p}{\partial \theta^i}\frac{\partial p}{\partial \theta^j} d\mathbf{x}$, then the GMM would reduce to separable one dimensional Gaussian integrals for which the closed-from solution exists. This translates to selecting a ϕ such that ϕ'' becomes a constant in (6). In [12], Havrda and Charvát introduced the notion of a α-order entropy using the convex function

$$\phi(p) = (\alpha - 1)^{-1}(p^\alpha - p), \alpha \neq 1 \, . \tag{8}$$

As $\lim_{\alpha \to 1} \phi(p)$, (8) goes to (7). We set $\alpha = 2$ which results in $\frac{1}{2}\phi'' = 1$. The one-half scaling factor does not impact the metric properties. The new metric is defined as

$$g^\alpha_{i,j} = \int_\chi (\frac{\partial p}{\partial \theta^i})(\frac{\partial p}{\partial \theta^j}) d\mathbf{x} \tag{9}$$

and we refer to it as the α-*order entropy metric tensor*.

The new metric can be used to find the distance between two shapes represented as $p(\mathbf{x}|\theta_1)$ and $p(\mathbf{x}|\theta_2)$. The geodesic can be obtained by solving

$$g^\alpha_{ki}\ddot{\theta}^i + \Gamma_{k,ij}\dot{\theta}^i\dot{\theta}^j = 0 \tag{10}$$

where $\Gamma_{k,ij} \stackrel{\text{def}}{=} \frac{1}{2}\left\{\frac{\partial g^\alpha_{ik}}{\partial \theta^j} + \frac{\partial g^\alpha_{kj}}{\partial \theta^i} - \frac{\partial g^\alpha_{ij}}{\partial \theta^k}\right\}$ is the Christoffel symbol of the first kind and standard Einstein summation convention is assumed. Following [10], we use gradient descent to find a local solution to this system of second order ODEs. However, unlike [10], we now have closed-form solutions to $g^\alpha_{i,j}$ and $\frac{\partial g^\alpha_{kj}}{\partial \theta^i}$. (We have elected to show the form of the metric tensor and its derivative for one particular combination of indices below and partially list the rest in the Appendix.) Here the θ-coordinates belong to the same landmark as indicated by $\lceil \frac{i}{2} \rceil = \lceil \frac{j}{2} \rceil$ and the i and j indices are equal.

$$\begin{array}{l} \text{If } \lceil \frac{i}{2} \rceil = \lceil \frac{j}{2} \rceil, i = j \Rightarrow g_{ij} = \frac{1}{8\pi\sigma^4 K^2} \\ \qquad\qquad\qquad\qquad\quad \frac{\partial g_{ij}}{\partial \theta^k} = 0. \end{array} \tag{11}$$

Please see the Appendix for more details regarding the metric tensor and Christoffel symbols.

4 Experimental Results and Analysis

In order to get a better understanding of the geodesics produced with the Fisher-Rao metric versus the new α-order entropy metric, we first computed the geodesics for a set of simple shapes. Figure 2 illustrates geodesics obtained from matching a four-landmark square to one that has been rotated 210° clockwise. The geodesics obtained

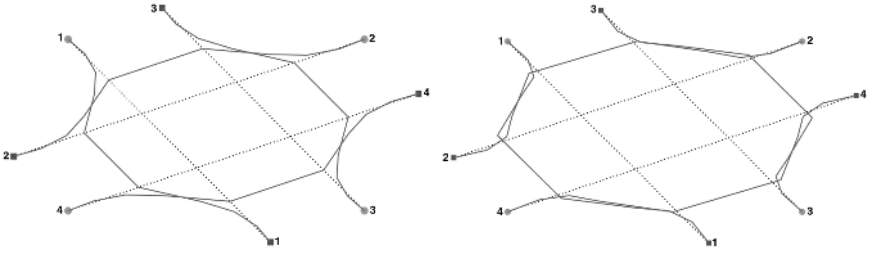


Fig. 2. Fisher information versus α-order entropy metric. The dashed line is the initialization and the solid line the final geodesic. The corresponding points are numerically labeled where circular landmarks are the starting shape and square landmarks the rotated shape. (a) Fisher information metric path is curved smoothly. (b) α-entropy metric path has sharp corners.

by the Fisher-Rao metric are smoothly curved, illustrating the hyperbolic nature of the manifold induced by the information matrix [14] whereas the α-order entropy metric displays sharper, abrupt variations. This is due to the fact that the Christoffel symbol values are zero for (i, j) pairs that belong to the same landmark.

For applications in medical imaging, we have evaluated both metrics on real data consisting of nine corpora callosa with 63-landmarks each as shown in Figure 3 in the Appendix. We also compared the results to the standard thin-plate spline (TPS) bending energy and studied the particular sensitivities of each metric. The results of pairwise matching of all nine shapes is listed in Table 1. For brevity, only the best and worst matching distance for each pair are listed. In instances where the TPS distance disagrees with other metrics, it assigns the worst values to shapes that differ largely due to "bending" transformations. On the other hand, both the Fisher-Rao and α-order entropy metric take into account how neighboring landmarks change and reflect sensitivities to both bending and stretching. This has good potential for computational medical applications such as computer-aided diagnosis and biological growth analysis where one is interested in measures that take into account multiple, local deformations. Though we have shown results on shapes that exhibit closed-contour topology, it is worth noting that this formulation has no inherent topological constraints on shape structure—thus enabling landmark analysis for anatomical forms with interior points or disjoint parts.

The α-order entropy metric provided huge computational benefits over the existing method utilizing the Fisher-Rao metric. The Fisher-Rao metric requires an extra $O(N^2)$ computation of the integral over $\mathbb{R}^2$ where we have assumed an N point discretization of the x- and y-axes. This computation must be repeated at each point along the evolving

geodesic and for every *pair* of landmarks. The derivatives of the metric tensor which are needed for geodesic computation require the same $O(N^2)$ computation for every landmark *triple* and at each point on the evolving geodesic. Since our new ϕ-entropy metric tensor and derivatives are in closed-form, this extra $O(N^2)$ computation is not required. Please note that the situation only worsens in 3D where $O(N^3)$ computations will be required for the Fisher-Rao metric (and derivatives) while our new metric (and derivatives) remain in closed-form. It remains to be seen if other closed-form information metrics can be derived which are meaningful in the shape matching context.

5 Conclusion

In this paper, we introduced a new Riemannian metric for landmark-based shape matching that combines both shape representation and deformation. This improved on the previous work of [10] where the Fisher-Rao metric was used to establish an intrinsic metric between landmark shapes. Given that our parameter space comes from Gaussian mixture models, the Fisher-Rao metric suffers serious computational inefficiencies as it is not possible to get closed-form solutions to the metric tensor or the Christoffel symbols. The new α-order entropy metric, with $\alpha = 2$, enables us to obtain closed-form solutions to the metric tensor and its derivatives and therefore alleviates this computational burden. Our new approach also illustrated the possibilities of using information metrics besides Fisher-Rao and their benefits. Test results on corpus callosum shapes show the applicability of the new metric to shape matching, providing discriminability similar to that of the Fisher-Rao metric—though admittedly these results are still in the anecdotal stage since we have not performed statistical comparisons on the computed shape geodesic distances. In addition, comparison to the popular thin-plate spline bending energy illustrates the sensitivities of both information based metrics to local deformations other than just bending. These metrics also do not suffer from topological constraints on the shape structure (thus enabling their applicability to a large class of medical imaging applications). Our future work will focus on extending this framework to incorporate diffeomorphic warping of the extrinsic space, intrinsically solving for correspondences between the shape landmarks and extensions to 3D shape matching.

Acknowledgements

This work is partially supported by NSF IIS-0307712 and NIH R01NS046812. We acknowledge helpful conversations with Chris Small and early reviews by Tariq Bakir and Eric Spellman.

References

1. Small, C.: The statistical theory of shape. Springer, New York, NY (1996)
2. Bookstein, F.L.: Principal warps: Thin-plate splines and the decomposition of deformations. IEEE Trans. Patt. Anal. Mach. Intell. **11** (1989) 567–585

3. Rohr, K., Stiehl, H., Sprengel, R., Buzug, T., Weese, J., Kuhn, M.: Landmark-based elastic registration using approximating thin-plate splines. IEEE Trans. on Medical Imaging **20** (2001) 526–534
4. Davies, R., Twining, C., Cootes, T., Taylor, C.: An information theoretic approach to statistical shape modelling. In: European Conference on Computer Vision (ECCV). Lecture Notes in Computer Science, LNCS 2351, Springer (2002) III: 3–20
5. Camion, V., Younes, L.: Geodesic interpolating splines. In: Energy Minimization Methods for Computer Vision and Pattern Recognition. Springer, New York (2001) 513–527
6. Joshi, S., Miller, M.: Landmark matching via large deformation diffeomorphisms. IEEE Trans. Image Processing **9** (2000) 1357–1370
7. Chui, H., Rangarajan, A.: A new point matching algorithm for non-rigid registration. Computer Vision and Image Understanding **89** (2003) 114–141
8. Guo, H., Rangarajan, A., Joshi, S., Younes, L.: Non-rigid registration of shapes via diffeomorphic point matching. ISBI 2004 (2004)
9. Siddiqi, K., Shokoufandeh, A., Dickinson, S.J., Zucker, S.W.: Shock graphs and shape matching. In: ICCV. (1998) 222–229
10. Peter, A., Rangarajan, A.: Shape matching using the Fisher-Rao Riemannian metric: Unifying shape representation and deformation. ISBI 2006 (2006)
11. Burbea, J., Rao, R.: Entropy differential metric, distance and divergence measures in probability spaces: A unified approach. Journal of Multivariate Analysis **12** (1982) 575–596
12. Havrda, M.E., Charvát, F.: Quantification method of classification processes: Concept of structural α-entropy. Kybernetica **3** (1967) 30–35
13. Rao, C.: Information and accuracy attainable in estimation of statistical parameters. Bulletin of the Calcutta Mathematical Society **37** (1945) 81–91
14. Costa, S., Santos, S., Strapasson, J.: Fisher information matrix and hyperbolic geometry. IEEE Information Theory Workshop (2005) 28–30

Appendix

Metric Tensor and Christoffel Symbol Components

Below we list the closed-form expressions for a few combinations of the metric tensor and its derivative.

$$\text{If } \lceil \tfrac{i}{2} \rceil = \lceil \tfrac{j}{2} \rceil, i \neq j:$$
$$g_{ij} = 0,$$
$$\frac{\partial g_{ij}}{\partial \theta^k} = 0$$

$$\text{If } \lceil \tfrac{i}{2} \rceil \neq \lceil \tfrac{j}{2} \rceil, i \text{ odd } j \text{ odd}:$$
$$g_{ij} = \frac{e^{-\frac{(\theta^i-\theta^j)^2}{4\sigma^2}} e^{-\frac{(\theta^{i+1}-\theta^{j+1})^2}{4\sigma^2}}}{16\pi^2\sigma^8K^2}\{\pi\sigma^2[2\sigma^2-(\theta^i)^2+2\theta^i\theta^j-(\theta^j)^2]\}$$
$$\text{for } k = i, j:$$
$$\frac{\partial g_{ij}}{\partial \theta^k} = -\frac{\pi e^{-\frac{(\theta^i-\theta^j)^2}{4\sigma^2}} e^{-\frac{(\theta^{i+1}-\theta^{j+1})^2}{4\sigma^2}}}{32\pi^2\sigma^8K^2}(\theta^i-\theta^j)[6\sigma^2-(\theta^i)^2+2\theta^i\theta^j-(\theta^j)^2]$$

$$i \text{ odd } j \text{ even}:$$
$$g_{ij} = -\frac{e^{-\frac{(\theta^i-\theta^{j-1})^2}{4\sigma^2}} e^{-\frac{(\theta^{i+1}-\theta^j)^2}{4\sigma^2}}}{16\pi^2\sigma^8K^2}\{\pi\sigma^2[(-\theta^{i+1}+\theta^j)(-\theta^i+\theta^{j-1})]\}$$
$$\text{for } k = i:$$
$$\frac{\partial g_{ij}}{\partial \theta^k} = \frac{\pi e^{-\frac{(\theta^i-\theta^{j-1})^2}{4\sigma^2}} e^{-\frac{(\theta^{i+1}-\theta^j)^2}{4\sigma^2}}}{32\pi^2\sigma^8K^2}(-\theta^{i+1}+\theta^j)[2\sigma^2-(\theta^i)^2+2\theta^i\theta^{j-1}-(\theta^{j-1})^2]$$

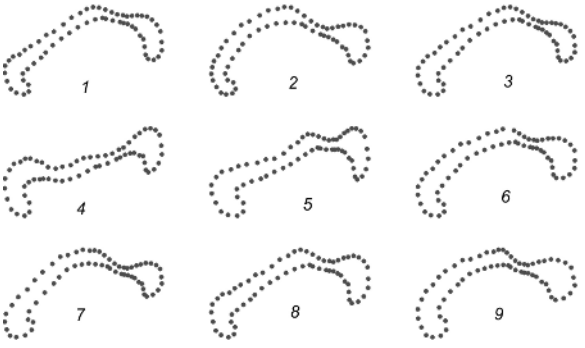


Fig. 3. Nine corpus callosum shapes used for pairwise matching, 63 landmarks per shape

Table 1. Pairwise shape distances. Column 2 and 3 are obtained by integrating the intrinsic geodesic using the Fisher-Rao and α-order entropy metric respectively. Column 4 is the bending energy of the TPS warp. Only the best and worst distances are shown are shown for each pair. For example, matching corpus callosum 1 against the remaining 8 shapes resulted in Fisher and the α-order metric assigning the shortest distance (i.e. best match) to shape 8 and worst distance to shape 4, whereas TPS assigned the best match to shape 3 and worst match to shape 7.

Pairs	Fisher-Rao (10^{-2})	α-Order Entropy (10^{-3})	TPS(10^{-2})
1 vs. 3	14.5905	2.6643	1.4664
1 vs. 4	87.0364	16.2904	10.5958
1 vs. 7	52.8142	10.9610	13.2774
1 vs. 8	3.0751	0.6217	3.0467
2 vs. 4	136.5600	23.8346	17.6227
2 vs. 6	15.8750	3.2368	5.5466
3 vs. 4	153.6000	28.0696	11.1349
3 vs. 6	6.2069	1.1633	2.7389
3 vs. 8	13.3270	2.2415	2.0639
4 vs. 5	59.5560	13.1824	12.8418
4 vs. 7	193.4700	37.7694	34.3238
5 vs. 7	78.6250	12.3373	13.8462
5 vs. 8	20.2640	3.4986	4.0704
6 vs. 7	22.2960	4.0647	5.7042
6 vs. 8	5.9597	1.1400	3.3339
7 vs. 8	40.0820	7.8548	6.8678
7 vs. 9	21.3106	3.9808	5.1912
8 vs. 9	23.9451	5.6155	9.2446

Feasibility of Patient Specific Aortic Blood Flow CFD Simulation

Johan Svensson[1,5], Roland Gårdhagen[1,5], Einar Heiberg[2], Tino Ebbers[3,5], Dan Loyd[1], Toste Länne[3,5], and Matts Karlsson[4,5]

[1] Department of Mechanical Engineering, Linköping University, Sweden, johsv@ikp.liu.se
[2] Department of Clinical Physiology, Lund University, Sweden
[3] Department of Medicine and Care, Linköping University, Sweden
[4] Department of Biomedical Engineering, Linköping University, Sweden
[5] Center for Medical Image Science and Visualization (CMIV), Linköping University, Sweden

Abstract. Patient specific modelling of the blood flow through the human aorta is performed using computational fluid dynamics (CFD) and magnetic resonance imaging (MRI). Velocity patterns are compared between computer simulations and measurements. The workflow includes several steps: MRI measurement to obtain both geometry and velocity, an automatic levelset segmentation followed by meshing of the geometrical model and CFD setup to perform the simulations follwed by the actual simulations. The computational results agree well with the measured data.

1 Introduction

The aorta is the primary blood vessel for transporting blood from the left heart to the systemic circulation. It has a very complex three-dimensional geometry including curving, branching and tapering. To be able to describe this very complex flow situation in detail in the normal human as well as for different pathological conditions in various diseases, a detailed description of the flow situation is crucial.

One common systemic disorder that can be studied with the use of detailed velocity information is atherosclerosis. Atherosclerosis is the main reason for many cardiovascular diseases [1]. However, atherosclerosis is not uniformly localized which indicates an influence from the blood velocity field on the genesis of atherosclerosis [2]. The near wall flow of the artery has thus been in focus, often in the form of wall shear stress (WSS) i.e. the frictional load on the artery wall. WSS have been intensively used to investigate the correlation with atherosclerosis all the way back to the first studies by [3] and [4]. Although there have been attempts to estimate WSS from direct flow measurements [5] the process is still inherently hampered with the need to make critical assumptions on the flow profile and three dimensional phase contrast magnetic resonance imaging (PC-MRI) velocity measurements lack the necessary resolution in the near

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 257–263, 2006.

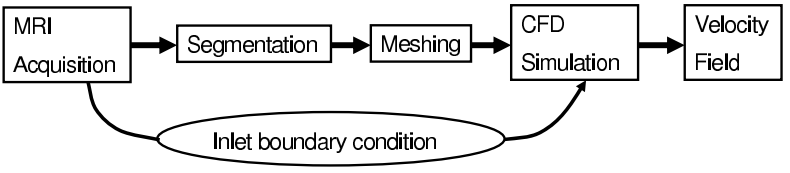


Fig. 1. Workflow overview

wall region. Thus, there are currently no non-invasive methods that can reliably be used to compute WSS directly to study or monitor the progress of the atherosclerosis.

To circumvent this CFD simulations are used to obtain a much more detailed view of the blood flow situation near the wall in the arteries. Using CFD techniques in the cardiovascular system have been made by numerous of research groups e.g. [6, 7, 8, 9, 10] and is constantly beeing refined and developed. However, creating a patient specific flow simulation model of a human aorta includes a wide range of engineering and medical disciplines. Some important parts are: state-of-the-art imaging techniques with sufficient resolution and quality, fast and accurate segmentation methods to gain a correct geometry and the very latest in CFD technology.

As a validation the flow simulation results are compared with 2D MRI velocity measurements. This kind of validation between CFD and measured MRI velocities have previously been performed by other groups before [11, 12] but studies conducted on patient specific models on large arteries in-vivo are rare [12, 13].

To be able to use the proposed approach in a clinical situation all the steps need to be highly automated and fast. The aim of this work is to demonstrate the feasibility to produce accurate patient specific flow simulations in the human aorta that can be used in a clinical situation.

2 Method

The workflow is outlined in Figure 1. Starting from the left we need MRI measurement to obtain both geometry and velocity, an automatic levelset segmentation followed by meshing (the geometrical model) and finally the CFD setup to perform the simulations. The method is described in cronological order.

2.1 MRI Acquisition

Magnetic Resonance Imaging was used to collect data from 2 male volunteers (20 denoted (I) and 25 denoted (II) year) using a 1.5 T MRI scanner (Philips Achieva, Philips Medical Systems, Best, the Netherlands). Geometrical information of the complete aorta was obtained within a breath hold using a 3D gadolinium-enhanced gradient-echo sequence (TE 1.6 ms, TR 5.3 ms, and flip angle 40, field of view 400x360x80 mm, acquisition matrix 400x207x80, SENSE factor 1.5). The 30 ml (0.5 mmol/ml) contrast bolus (Omniscan, Amersham Health, Oslo,

Norway) was injected at 2.0 ml/s. Randomly segmented central k-space ordering (CENTRA) was used. The three-dimensional volume data was reconstructed to a resolution of 0.78x0.78x1.00 mm.

Time-resolved information of the aortic flow velocities were obtained by performing a through-plane 2D velocity MRI acquisition (echo time 2.3 ms, repetition time 3.9 ms, flip angle 15, velocity encoding range 2.00 m/s) placed supracoronary perpendicular to the flow direction. The acquisition was performed during a breath hold using SENSE factor 2.0 and retrospective cardiac gating to a vectorcardiogram (VCG). The 10 mm thick slice was acquired with a field of view of 350x297 mm, acquisition matrix 144x122, and a temporal resolution of 31.2 ms (I) or 39 ms (II). The acquired data were reconstructed to 40 timeframes per heart cycle with a spatial resolution of 1.37x1.37 mm. The velocity data were corrected for effects of concomitant gradient fields and eddy currents.

2.2 Segmentation

The purpose of the segmentation is to create a detailed three-dimensional geometric description of the aorta that can be used for meshing. The accuracy of the segmentation is crucial for the final CFD simulation result [14, 15, 16]. We used a fast level set algorithm originally proposed by [17]. In a level set approach one starts with a small seed point and an implicit surface is allowed to expand outwards [18]. The expansion speed is determined by a speed image and the local curvature. The speed image was calculated as:

$$\mathbf{I}_{Speed}(\boldsymbol{x}) = \alpha_0 \mathbf{I}(\boldsymbol{x}) - \alpha_1 \mathbf{I}_{Edge}(\boldsymbol{x}) - \bar{I}_{Seed} + \alpha_2 \tag{1}$$

where $\mathbf{I}_{Speed}$ is speed image, $\boldsymbol{x}$ indicates spatial dependency, α_n are user adjusted coefficients, $\mathbf{I}$ is the MRI image, $\bar{I}_{Seed}$ is mean intensity of the seed points, and $\mathbf{I}_{Edge}$ is an edge image created by sobel filtering. The value of the coefficients were not very critical. They were roughly set to $\alpha_0 = 7$, $\alpha_1 = 8$, and $\alpha_2 = 0$.

The expansion speed were calculated as:

$$F(\boldsymbol{x}) = \mathbf{I}_{Speed}(\boldsymbol{x}) + \beta \boldsymbol{\kappa}(\boldsymbol{x}) \tag{2}$$

where F is the local expansion speed, $\mathbf{I}_{Speed}$ is the speed image, β is a fix coefficient set to 0.45, and κ finally is the local curvature of the level set surface.

By using a virtual 3D-pen with variable thickness the user could place seed points and at some places force the expansion velocity to zero. Manual intervention were typically needed to avoid the heart and pulmonary artery. The total time for the segmentation was around one hour per data set, but might be significantly shortened by improved user interface and practice. The used levelset algorithm is essentially a fast marching algorithm with an approximation to calculate the local curvature. This allows the algorithm to be comparable in speed with fast marching algorithms but still use curvature to produce a smoother result. The algorithm is binary so the final result is a binary mask of the aorta.

The algorithm was implemented into a cardiac image analysis software package (`http://segment.heiberg.se/`).

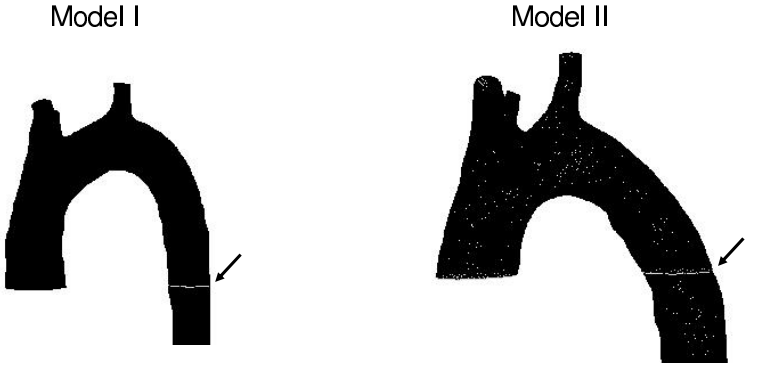


Fig. 2. Meshes of the aorta model I (left) and Model II (right). Location of the comparision cross-sections are marked by the arrows.

2.3 Meshing

The first step is to smooth the binary result from the level set algorithm. This was done by applying a 3D Gaussian smoothing filter. The radius of the filter was set to 2 mm where the distance denotes the point where the energy of the filter has dropped to $1/e$. An isosurface was created using the marching cubes algorithm from the smoothed result of the binary mask with the iso level of 0.5. The isosurface was then converted to a sterolithography format (`.STL`). This file could the be imported by the mesh generator ICEM 10.0. Due to the smoothing procedure the geometry at the very proximal parts of inlets and outlets were changed so these parts were cropped of to ensure a correct geometry. An unstructured mesh was applied to the geometry, this choice was made because of the complexity in the geometry. A more quad element based mesh can be used but requiring more manual intervention. The meshes created are outlined in Figure 2.

Because that interesting areas for atherosclerosis are near the artery wall the mesh near the wall was refined by inserting four prism layers, see Figure 3.

2.4 CFD Simulation

The CFD simulations are performed with the commercial software Fluent version 6.1.18, which uses a finite volume method. The governing equations are the Navier-Stokes equations consisting of the continuity equation, Equation 3, and the momentum equations, Equation 4.

$$\oint \rho \mathbf{V} \cdot d\mathbf{A} = 0 \tag{3}$$

$$\oint \mathbf{V}\mathbf{V} \cdot d\mathbf{A} = -\oint p\mathbf{I} \cdot d\mathbf{A} + \oint \tau \cdot d\mathbf{A} \tag{4}$$

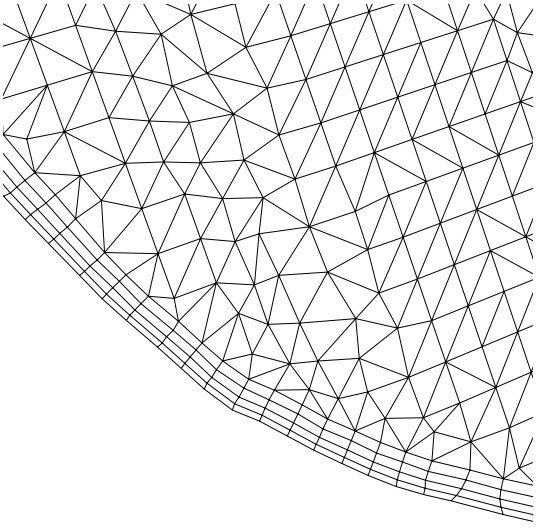


Fig. 3. Prism layers at the inlet in one of the models

where $\mathbf{V}$ is the velocity vector, ρ is the density, p is the pressure, $\mathbf{I}$ is the identity matrix, and τ is the stress tensor. The density of the fluid was set to 1060 kg/m^3 and the viscosity to 0.00345 Ns/m^2.

The velocity profile is measured with 2D MRI in a cross section in the ascending aorta and used as the inflow boundary condition. The outflow boundary conditions ensure conservation of mass. 15% of the flow is assumed to leave the model through the first branch (brachiosephalic trunk), 5% through the second branch (left common carotid), 10% through the third branch (left subclavian artery), and 70% exits through the abdominal outflow. The arterial wall in the model is treated as rigid.

3 Results and Discussion

The results of the CFD simulation were evaluated in a cross-section in the decending aorta where we both had CFD velocity field and 2D MRI velocities. The results used are at three different times when the flow accellerates (t_1) near the peak velocity (t_2) and at the deccelaration (t_3) phase. Both measured and CFD simulated velocities are seen in Figure 4.

From MRI images through segmentation and meshing procedure it takes less then two hours and the result is a suffiently accurate patient specific mesh ready for CFD simulations.

A good agreement is found between the stationary CFD simulation results to the PC-MRI measured velocities (Figure 4). Both flow profile and the velocity magnitude shows good agreement. There are difference as well that may depend on the chosen outflow in the aortic arch branches.

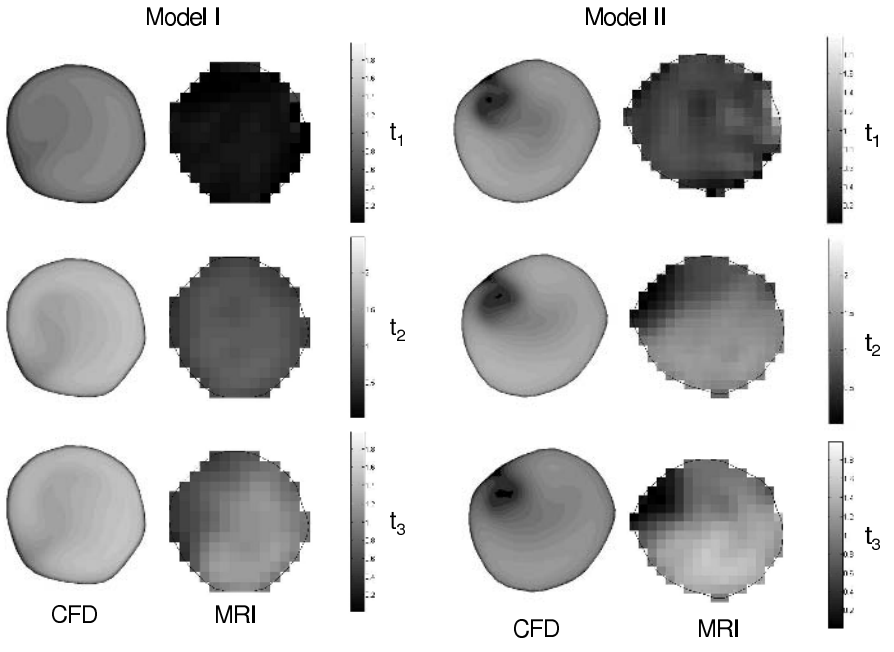


Fig. 4. Velocity magnitude in a decending aorta cross-section from both PC-MRI measurement and CFD simulations. Time position t_1, t_2 and t_3 correspond to accelerating flow, peak velocity and deaccelerating flow.

Using the described setup it is fully feasible with todays technology to scan a patient in the afternoon, and have an accurate flow simulation the following morning.

Future research will focus on further improvement to the segmentation process to further minimize the need for user interaction. One possible improvement would be to use template based method for segmentation. We also plan to perform time resolved CFD simulation, and to do larger studies involving larger number of subjects.

References

1. Berenson, G., Srinivas, S., Bao, W., Newman III, W., Tracy, R., Wattingney, W.: Association between multiple cardiovascular risk factors and atherosclerosis in childeren and young adults. The New England Journal of Medicine **338** (1998) 1650–1656
2. Humprey, J.: Cardiovascular Solid Mechanics. Springer-Verlag (2002)
3. Caro, C., Fitz-Gerald, J., Schroter, R.: Arterial wall shear: Observation, correlation and proposal of a shear dependent mass transfer mechanism for atherogenisis. In: Procedings of Royal Society in London. Volume B:177. (1971) 109–159
4. Fry, D.: Acute vascular endothelial changes associated with increased blood velocity gradient. Circular Research **22** (1968) 165–197

5. Wu, S., Ringgaard, S., Oyre, S., Hansen, M., Rasmus, S., Pedersen, E.: Wall shear rates differ between the normal carotid, femoral, and bracial arteries: An in vivo MRI study. Journal of Magnetic Resonance Imaging (2004) 188–193
6. Quarteroni, A., Tuveri, M., Veneziani, A.: Computational vascular fluid dynamics: problems, models and methods. Computing and Visualization in Science **2** (2000) 163–197
7. Cebral, J., Castro, M., Soto, O., Löhner, R., Alperin, N.: Blood-flow models of the circle of willis from magnetic resonance data. Journal of Engineering Mathematics **47** (2003) 369–386
8. Finol, E., Amon, C.: Flow-induced wall shear stress in abdominal aortic aneurysms: Part i - steady flow hemodynamics. Computer Methods in Biomechanics and Biomedical Engineering **5** (2002) 309–318
9. Finol, E., Amon, C.: Flow-induced wall shear stress in abdominal aortic aneurysms: Part ii - pulsatile flow hemodynamics. Computer Methods in Biomechanics and Biomedical Engineering **5** (2002) 319–328
10. Steinman, D., Thomas, J., Ladak, H., Milner, J., Rutt, B., Spence, J.: Reconstruction of carotid bifurcation hemodynamics and wall thickness using computational fluid dynamics and mri. Magnetic Resonance in Medicine **47** (2002) 149–159
11. Glor, F., Westenberg, J., Vierendeels, J., Danilouchkine, M., Verdonck, P.: Validation of the coupling of magnetic resonance imaging velocity measurement with computational fluid dynamics in an U bend. Artificial Organs **27** (2002) 622–635
12. Ku, J., Elkins, C., Taylor, C.: Comparison of CFD and MRI flow and velocities in an in vitro large artery bypass graft model. Annals of Biomedical engineering **33** (2005) 257–269
13. Leuprecht, A., Perktold, K., Kozerke, S., Boesiger, P.: Combined cfd and mri study of blood flow in a human ascending aorta model. Biorheology **39** (2002) 425–429
14. Friedman, M., Deters, O., Mark, F., Bargeron, C., Hutchins, G.: Arterial geometry affects hemodynamics. A potential risk factor for atherosclerosis. Atherosclerosis **46** (1983) 225–231
15. Svensson, J., Gårdhagen, R., Karlsson, M.: Geometrical considerations in patient specific models of a human aorta with stenosis and aneurysm. In: Computational Fluid Dynamics 2004 Proceedings of the Third International Conference on Computational Fluid Dynamics, ICCFD3, Toronto, 12-16 July 2004, ISBN: 3-540-31800-3. (2004)
16. Svensson, J., Gårdhagen, R., Karlsson, M.: Assesment of geometrical influence on wss estimation in the human aorta. Accepted for publication in: WSEAS Transactions on Fluid Mechanics (2006)
17. Nilsson, B., Heyden, A.: A fast algorithm for level set-like active contours. Pattern Recognition Letters **24** (2003) 1331–1337
18. Sethian, J.A.: Level set methods and fast marching methods. Cambridge University Press, Cambridge, USA (1999) ISBN 0 521 64557 3.

A Model Based Approach for Multi-lead ECG Array Layout Selection

Christoph Hintermüller, Michael Seger, Bernhard Pfeifer,
Gerald Fischer, and Bernhard Tilg

Institute for Biomedical Engineering,
University for Health Sciences, Medical Informatics and Technology,
Eduard Wallnöfer Zentrum I, 6060 Hall in Tyrol, Austria
Telephone: 0043/50/8648/3820, Fax: 0043/50/8648/3850

Abstract. In this study an approach for testing electrode array schemes with respect to their ability to improve the resolution of methods for activation time imaging is proposed. First local linear dependency maps are computed using a virtual array method. These maps depict the torso surface areas where the body surface potential is most sensitive to changes in the transmembrane potential. The optimal number and position of the electrodes within the sensitive body surface regions was selected by constructing effort gain (EG) plots. Such a plot depicts the relative attainable rank of the leadfield matrix in relation to the increase in number of electrodes required to build the electrode array.

From the sensitivity maps it was found that the BSP is most sensitive to changes in TMP on the upper left frontal and dorsal body surface. The EG analysis revealed that the optimal array meeting clinical requirements and improving the resolution of activation time imaging consists of 125 electrodes.

1 Introduction

Here methods for selecting an electrode array scheme maximizing the information content in body surface potential recordings are proposed.

Observations upon linear dependency between electrodes are used to locate areas of high information content in the BSP patterns by calculating so called local linear dependency (LLD) maps. The volume conductor model required for the LLD analysis was extracted from patient individual magnetic resonance imaging (MRI) scans, recorded prior to the clinical treatment the patient had to undergo. As depicted in Fig. 1 it basically consists of four different compartments, the chest surface, the left and right lungs, the ventricular and atrial myocardium and the associated bloodmasses. Further, it contains information upon the electric conductivity of the different compartments and the positions of the electrodes the BSP is recorded with. Especially locating the posterior electrodes is a rather tedious procedure as their position is reconstructed from the MRI scans by locating and identifying vitamin E marker capsules. Whereas the positions of the frontal ones can be recorded using a digital position tracking

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 264–271, 2006.

system such as FASTRACK (Polhemus Inc, Colchester, USA). Recording the posterior electrodes with the FASTRACK system would require the patient to sit upright which usually is not possible. Further, the proper electrical contact of all electrodes to the body surface of the patient has to be considered critical. In case of improper contact the signal to noise ratio (SNR) is significantly lower than the required level (in our case >40 dB).

The BSP recordings were used to estimate the electric activation for the heart by computing the activation time as described in [1, 2, 3, 4, 5]. The information upon the electric sources of the heart was noninvasively obtained by Recording the BSP. The BSP is used in combination with an individual volume conductor model to solve the underlying inverse problem. Before solving this inverse problem, the type of information which is determined in the inverse problem has to be specified. For detecting the earliest activation of the heart an AT imaging method will be the best choice. In any case the changes in the TMP related to the activation of the heart cause changes in the BSP. The corresponding solutions are sensitive to the recording location of the BSP.This sensitivity is caused by varying linear dependencies in the equation system obtained for the forward electrocardiographic problem. In case of high linear dependency the BSP contains less additional information while BSP recorded from locations with low linear dependency contains more independent information about cardiac electric activity.

2 Methods

For solving the electrocardiographic forward and inverse problem a patient individual volume conductor model is used. Such a model the patients torso, lungs, ventricles, atria and associated blood masses and the location of the electrodes. The different compartments have been segmented using methods and tools described in [6]. The volume conductor includes the compartment conductivities κ and κ_{H} for the heart. Using discrete boundaries surfaces with the interpolation coefficients n_i, $n_{j,\mathrm{H}}$ and the potentials $V_i(t)$, $\varphi_{\mathrm{m},j}(t)$ and applying the Wilson Terminal formulation as done by [7] the relation between the TMP φ_{m} on the heart surface and the BSP V measured by the electrodes can be expressed as

$$V_{\mathrm{e}}(t) = \mathbf{IN}\varphi_{\mathrm{m}}(t) = \mathbf{L}\varphi_{\mathrm{m}}(t) \ . \tag{1}$$

The leadfield matrix $\mathbf{L}$ in (1) consists of the interpolation matrix $\mathbf{I}$ interpolating the discrete BSP to gain the electrode potential and the node field matrix $\mathbf{N}$ describing the patients volume conductor.

2.1 Local Linear Dependency

In order to identify the sensitive areas a virtual array (VA) method which is similar to the virtual displacement method was applied. The distance Δd between the virtual electrodes and the central one located at the point of investigation P_{ci} is not infinitesimally small, which would be the case for the virtual displacement

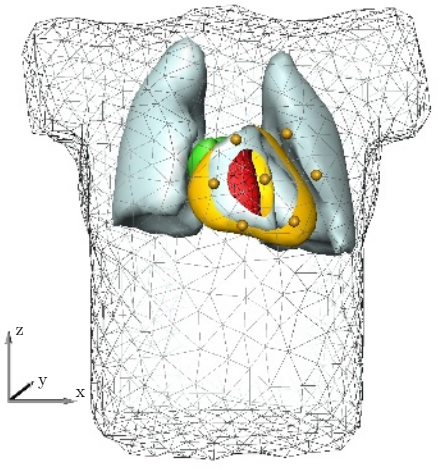


Fig. 1. Volume conductor model consisting of the atria, the ventricles, the corresponding blood masses, the lungs and the chest. A seven lead virtual array, centered at a node of the body surface mesh, is depicted. The right ventricle is cut open to show the corresponding blood mass.

method. Figure 1 shows such a VA with P_{ci} equal to a node of the triangular body surface mesh.

From the leadfield matrix $\mathbf{L}_{\mathrm{virt}}$ of a VA a LLD measure l with

$$l = \frac{s_{\max}}{s_{\min}} \tag{2}$$

was computed. In (2) $s_{\min}$ and $s_{\max}$ are the smallest and largest singular value of $\mathbf{L}_{\mathrm{virt}}$. The singular values s_i are obtained by singular value decomposition (SVD) of $\mathbf{L}$ as follows:

$$\mathbf{L} = \mathbf{U}\mathbf{\Sigma}\mathbf{Q}^{\mathrm{T}} . \tag{3}$$

The rectangular matrix $\mathbf{U}$ contains the eigenvectors spanning the orthonormal basis of the electrode (row) space. The diagonal matrix $\mathbf{\Sigma}$ contains the singular values and $\mathbf{Q}$ contains the eigenvectors spanning the orthonormal basis of the source (column) space.

2.2 Attainable Rank

In addition to the VA based LLD maps the singular values s_i for the leadfield matrix $\mathbf{L}$ of an electrode array were computed. According to (1) and (3) the TMP $\varphi_{\mathrm{m}}(t)$ related to source space $\mathbf{Q}$ is mapped by $\mathbf{L}$ onto the BSP in electrode space $\mathbf{U}$. Thus, $\varphi_{\mathrm{m}}(t)$ can be expressed as a linear combination of the source eigenvectors q_i spanning $\mathbf{Q}$ as

$$\varphi_{\mathrm{m}}(t) = \Delta\varphi_{\mathrm{m}} \sum_i \alpha_i(t) q_i \ , \tag{4}$$

with the action potential amplitude $\Delta\varphi_{\mathrm{m}}$. The time dependent coefficients $\alpha_i(t)$ map q_i onto $\varphi_{\mathrm{m}}(t)$. They range from $\alpha_i(t) = 0$ corresponding to a total suppression of the spatial frequencies contained in q_i and $\alpha_i(t) = 1$ fully mapping q_i onto $\varphi_{\mathrm{m}}(t)$.

For estimating the attainable rank R of $\mathbf{L}$ a "detector" or "receiver criterion" based upon the following assumptions was applied:

1. The source distribution $\varphi_{\mathrm{m}}(t)$ is equal to $\varphi_{\mathrm{m}}(t) = \varphi_{\mathrm{m}} = \text{const}$ for all time instants t reducing (4) to

$$\varphi_{\mathrm{m}} = \Delta\varphi_m \sum_i \alpha_i q_i \ . \tag{5}$$

2. The source space $\mathbf{Q}$ is successively mapped onto φ_{m} by setting all $\alpha_i = 0$ in (4) except one $\alpha_i = \alpha_j = 1$. The index j is used to select one specific $q_{i=j}$. It ranges from $j = 1$ to the number of singular values s_i of $\mathbf{L}$.

Applying this assumptions to (1) the norm $\|V_{\mathrm{e}}(j)\|$ of the corresponding BSP vectors $V_{\mathrm{e}}(j)$ results in

$$\|V_e(j)\| = \|\mathbf{L}\Delta\varphi_{\mathrm{m}} q_{i=j}\| = s_{i=j}\Delta\varphi_m \ . \tag{6}$$

A detector in general is characterized by its measurement range, the number of leads Z the norm $\|\sigma\| = \sqrt{Z}\sigma_l$ of the noise vector σ containing the noise levels σ_l for all leads. The detector criterion uses $\|V_{\mathrm{e}}(j)\|$ and $\|\sigma\|$ to classify all $V_{\mathrm{e}}(j)$ with respect to their information content. It separates vectors $V_{\mathrm{e}}(j)$ resulting from the mapping of $\varphi_{\mathrm{m}}(j)$ to the target space $\mathbf{U}$ of $\mathbf{L}$ by (1) from vectors $V_{\mathrm{e}}(j) = \sigma$. The attainable rank R of $\mathbf{L}$ can be found by comparing all $\|V_{\mathrm{e}}(j)\|$ gained by (6) with $\|\sigma\|$ applying the following condition

$$s_{i=j+1}\Delta\varphi_{\mathrm{m}} \leq \|\sigma\| < s_{i=j}\Delta\varphi_{\mathrm{m}} \ . \tag{7}$$

The index j for which (7) holds is equal to the attainable rank R of $\mathbf{L}$.

2.3 Effort, Gain Analysis

One possibility to compare two electrode array layouts S_0 and S_i is to compare their attainable ranks R_0, R_i and the number of electrodes Z_0, Z_i used to build them. This was done by computing measures for the gain in rank r_i and the effort z_i when converting S_0 into S_i by changing position and number of electrodes. The measures z_i and r_i are computed in terms of Z and R as follows:

$$r_i = \frac{R_i}{R_0}, \quad z_i = \frac{Z_i}{Z_0} \ . \tag{8}$$

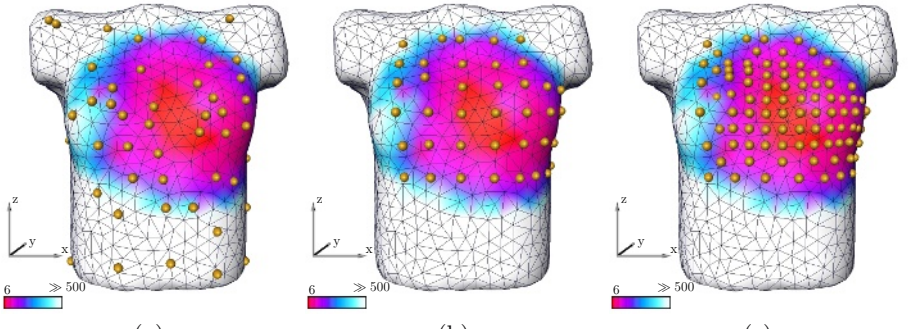
When scheme S_0 is compared to several S_i than the corresponding measures r_i and z_i can be used to construct a so called effort-gain (EG) plot. Such a plot depicts the relative increase in R with respect to the relative changes in Z. In general the EG-plot can be used to compare the arrays S_i to each other and select the optimal layout.

3 Results

3.1 Sensitive Body Surface Regions

Volume conductor models for eleven patients (9 male and 2 female) having different body shape, age (between 19 and 57 years) and size were generated. Each model was segmented from MRI scans recorded prior to the electrophysiological intervention. Based on these models the patient specific LLD maps for the atria, the ventricles and the whole heart were computed. Figure 2 depicts the anterior part of these LLD maps for an individual patient. By comparing the LLD maps for all 11 patients it was found that the sensitive areas expressing low l are located left frontal and dorsal between the shoulder blades. In average, the sensitive regions (gray area, Figure 2) were somehow L-shaped and showed a frontal size of $\sim 20 \times 30$ cm to $\sim 25 \times 35$ cm and $\sim 15 \times 20$ cm to $\sim 20 \times 25$ cm on the back.

In a first step the electrodes were arranged in two regular spaced L-shaped patches. In order to ensure that the sensitive regions for the atria, ventricles and the whole heart are covered best, a horizontal and vertical spacing of the electrodes equal to 4 cm (Fig. 2) was selected. The aim was to compare the array prototype of scheme S_B to an array scheme S_A, initially developed at the University of Amsterdam [8]. The number of electrodes for S_B was selected to be the same as for S_A. Due to the difficulties to extract the posterior electrode positions, as mentioned in section 1, the number of posterior electrodes had to be reduced form 21 to 11 for the layout of scheme S_A. Thus, S_B was compared to the modified array scheme S_A in the following addressed as $S_{A,62}$ instead of the original scheme S_A proposed by [8].



(a) (b) (c)

Fig. 2. The anterior linear dependency (LLD) map for an individual patient and its coverage by three different electrode arrays (spheres) (a), (b), (c). The gray surface indicates the frontal sensitive area (a) shows an idealized version $S_{A,62}$ of scheme S_A. As it can be observed it covers the anterior region poorly. The other two arrays $S_{B,62}$ (b) and $S_{B,125S}$ (c) expose a better coverage. Both are based upon the horizontal and vertical regular spaced layout introduced by scheme S_B.

In detail the initial prototype $S_{B,62}$ of S_B consisted of 62 Electrodes (41 anterior and 21 posterior). To perform the comparison of S_B with S_A the leadfield matrices $\mathbf{L}_{S_B}$ and $\mathbf{L}_{S_A}$ for the atria, the ventricles and the whole heart were computed. The corresponding ranks R were estimated using the detector criterion specified by (7). In addition, S_B and S_A were compared to a variant $S_{A,59}$ of scheme S_A which was recorded during the clinical procedure [5] the entire patient had to undergo.

Further, the figures for effort m and gain r were calculated. iFor all 11 patients the arrays $S_{B,62}$ and $S_{B,125S}$ exposed larger values for r than $S_{A,62}$ (eg: $r = 1$ for $S_{A,62}$, $r = 1.2$ for $S_{B,62}$ and $r = 1.76$ for $S_{B,125S}$). The arrays differ in their size, the body surface area they cover and the horizontal and vertical spacing of the electrodes. The scheme $S_{A,62}$ covers the whole body whereas $S_{B,62}$ only covers the sensitive areas. As it is suggested that a compact array is more applicable than a large one it may be concluded that $S_{B,62}$ is more suitable for our purpose than scheme $S_{A,62}$.

In case of the atria the values for the gain are $r = 1.2$ for $S_{B,62}$ and $S_{B,125S}$ and $r = 1$ for $S_{A,62}$ and $S_{A,59}$. The numbers obtained for the whole heart are similar to the ones for the ventricles, being $r = 1.07$ for $S_{B,62}$, $r = 1.64$ for $S_{B,125S}$ and $r = 1$ for $S_{A,62}$ and $S_{A,59}$.

3.2 Selection of Optimal Electrode Positions

Based on these findings and the suggestions given in [9] upon the clinical applicability of the L-shaped scheme S_B the search for the ideal number and density of electrodes was limited to L-shaped arrays derived from scheme S_B. As a consequence two groups of electrode arrays based on scheme S_B generated. The

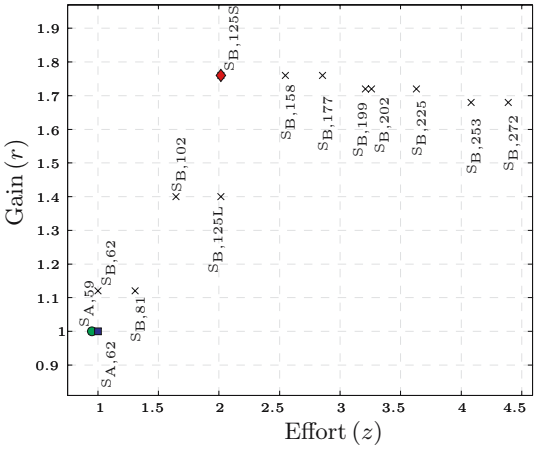


Fig. 3. Effort-gain-plot for one patient. The circular marker indicates the values for scheme $S_{A,59}$ recorded in the catheter lab. Its idealized version $S_{A,62}$ is depicted by the square marker and the diamond stands for $S_{B,62}$, $S_{B,125S}$ which we think perform best. In case of an atrial source $S_{B,62}$ would be sufficient whereas in case of the ventricles $S_{B,125S}$ should be the preferred choice.

first exposed the same shape and size as $S_{B,62}$ and the second was gained by expanding the size of the two L shaped parts. This enlargement was achieved by adding an additional row and column at a horizontal and vertical spacing of 4 cm to $S_{B,62}$ resulting in the 81 lead array $S_{B,81}$. For both sets further arrays were gained by subdividing their horizontal and vertical spacing.

For each of the 11 patients the models for the 14 arrays were generated and the corresponding rank R, gain r, effort m and the norm $\|\sigma\|$ of the detector noise vector σ were computed. The measures for m and r were used to generate patient individual EG plots, Fig. 3. Further the optimal array scheme was estimated. Figure 3 depicts the EG plots for the ventricles. When repeating the EG analysis for the atria, the ventricles and the whole heart, it was found that the 125 lead array scheme labeled $S_{B,125S}$ meet most of our requirements.For the atria the 62 lead array labeled $S_{B,62}$ would be sufficient.

4 Discussion

The results gained by the VA based LLD analysis lead to the assumption that it is possible to design an electrode array meeting the requirements given in [9]. This was mainly imposed by the fact that for all patients most information about the electrical excitation of the heart can be recorded from body surface regions located near to the heart. These regions, as depicted in Fig. 2, are located central and left on the frontal and dorsal body surface. The shape of the sensitive body surface regions can be roughly approximated by an L-shape. Thus, it was decided to use two L-shaped pieces to construct an electrode array meeting the clinical requirements. For the application of the frontal part it is suggest to use the six standard ECG leads V_1 to V_6 as landmarks. And the posterior array may be aligned along the spinal cord touching the tip of the left shoulder blade with its horizontal part.

By overlaying the LLD maps with models of the electrode scheme S_A as done in Fig. 2(a), it can be found that $S_{A,62}$ and $S_{A,59}$ do not cover the sensitive areas very well. Both expose an irregular electrode spacing and cover the patients body completely. This causes the poor coverage of the sensitive areas (colors) as can be seen in Fig. 2(a). Especially the low number of posterior electrodes, is enforced by the need to reconstruct their positions from MRI scans.

As mentioned in sec. 1 it was necessary to modify the scheme $S_{A,62}$ to be able to properly locate and identify the posterior electrodes. A similar identification and location process will be necessary for the posterior electrodes of $S_{B,62}$ and $S_{B,125S}$ too. Although they contain a lot more dorsal electrodes than $S_{A,62}$, it may be assumed that the regular horizontal and vertical spacing supports the following approach. The MRI scans of the patient are used to locate 2 - 7 characteristic electrodes. These electrodes may mark the edges of electrode array shape or some anatomic landmarks. The position of the remaining electrodes may then be reconstructed by matching the model of the posterior array with these landmarks. Even further, this two step approach could be used to locate the frontal electrodes too, instead of the FASTRACK (Polhemus Colchester Inc, USA) locator system.

References

1. Greensite, F.: Remote reconstruction of confined wavefront propagation. Inverse Problems **11** (1995) 361–370
2. Greensite, F.: Cardiac electromagnetic imaging as an inverse problem. Electromagnetics **21** (2001) 559–557
3. Huiskamp, G., Greensite, F.: A new method for myocardial activation imaging. IEEE Transactions on Biomedical Engineering **44** (1997) 433–446
4. Cuppen, J., Van Oosterom, A.: Model studies with inversely calculated isochrones of ventricular depolarization. IEEE Transactions on Biomedical Engineering **31** (1984) 652–659
5. Modre, R., Tilg, B., Fischer, G., Wach, P.: Noninvasive myocardial activation time imaging: A novel inverse algorithm applied to clinical ecg mapping data. IEEE Transactions on Biomedical Engineering **49** (2002) 1153–1161
6. Pfeifer, B., Seger, M., Hintermüller, C., Modre, R., Hanser, F., Tilg, B.: Patient-specific volume conductor modeling for noninvasive imaging of cardiac electrophysiology. IEEE Transactions on Medical Imaging (2005) (submitted).
7. Fischer, G., Tilg, B., Modre, R., Hanser, F., Messnarz, B., Wach, P.: On modeling the wilson terminal in the boundary and finite element method. IEEE Transactions on Biomedical Engineering **49** (2000) 217–224
8. SippensGroenewegen, A., Peeters, H.A.P., Jessurun, E.R., Linnenbank, A.C., de Medina, E.O.R., Lesh, M.D., van Hemel, N.M.: Body surface mapping during pacing at multiple sites in the human atrium. Circulation **1997** (1998) 369–380
9. Hintermüller, C., Fischer, G., Seger, M., Pfeifer, B., Hanser, F., Modre, R., Tilg, B.: Multi-lead ecg electrode array for clinical application of electrocardiographic inverse problem. In: Proceedings of the 26th Annual International Conference of the IEEE EMBS. (2004) 1941–1944

Simulation of Acquisition Artefacts in MR Scans: Effects on Automatic Measures of Brain Atrophy

Oscar Camara-Rey[1], Beatrix I. Sneller[1],Gerard R. Ridgway[1], Ellen Garde[2], Nick C. Fox[2], and Derek L. G. Hill[1]

[1] Center of Medical Image Computing, University College of London, UK
o.camara-rey@ucl.ac.uk
[2] Dementia Research Centre, Institute of Neurology,
University College Of London, UK

Abstract. Automatic algorithms in conjunction with longitudinal MR brain images can be used to measure cerebral atrophy, which is particularly pronounced in several types of dementia. An atrophy simulation technique has been devised to facilitate validation of these algorithms. To make this model of atrophy more realistic we simulate acquisition artefacts which are common problems in dementia imaging: motion (both step and periodic motion) and pulsatile flow artefact. Artefacts were simulated by combining different portions of k-space from various modified image. The original images were 7 MR scans of healthy elderly controls, each of which had two levels of simulated atrophy. We investigate the effect of the simulated acquisition artefacts in atrophy measurements provided by an automatic technique, SIENA.

1 Introduction

Structural magnetic resonance imaging (MRI) can be used to estimate the progression of neurodegenerative disease since brain atrophy is a measurable feature of the most common forms of dementia such as Alzheimer's disease (AD). Several computerized techniques [1-4] for the quantification of cerebral atrophy in MR scans are available in the literature. Nevertheless, there is a lack of consensus on the most appropriate method to be used since standardized test datasets with known atrophy rates are not available.

We have described [5] a method based on the combination of expert knowledge of clinical changes in brain with finite-element methods, providing valuable ground-truth data for the objective validation of atrophy measurement techniques. This technique, however, did not take into account the common presence of artefacts in the MR scans due to imperfections in the acquisition, which may have a strong impact on the performance of the subsequent atrophy quantification techniques. For instance, movement during the acquisition of the edge of k space can lead to a ringing artefact from the scalp [6], global movement during the central portion of k-space would

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 272 – 280, 2006.

generate obvious ghosts and blur image intensities, or flow artefacts may appear in the medial temporal lobes due to the pulsatile blood flow in the carotid arteries. All these artefacts have the potential to severely degrade quantitative MR-based analysis such as measurement of atrophy.

The main goal of this paper is to present the techniques we are developing to simulate acquisition artefacts in order to make the generated ground-truth data more realistic. Furthermore, these simulations will allow the assessment of post-processing artefact correction techniques [6, 7, 8] and their incorporation in atrophy measurement methods.

One of the most common causes of acquisition artefacts is subject motion. Rotational motion artefacts were simulated by Blumenthal et al. [9] to study their effects on a GM quantification technique and by Preboske et al. [10] to investigate their impact on a semi-automatic atrophy measurement method, the boundary shift integral [1]. Nevertheless, in some particular brain regions, other factors can be as important as subject motion. For instance, the main cause of image degradation in the temporal lobe, a region with a crucial role in Alzheimer's disease, is pulsatile flow artefact from carotid arteries. In clinical studies related to dementia, out of 837 subjects, we have visually identified 105 (12.5%) of the subjects' images with artefacts located in the temporal lobes, which can be classified according to their cause: 68 (64.8%) with flow artefact; 16 (15.2%) with motion-related artefact; 4 (3.8%) with susceptibility artefacts; and 17 (16.2%) images with artefacts induced by other factors.

Therefore, we have developed techniques to simulate the two most significant causes of acquisition artefacts, a rotational motion during the scan, and a pulsatile motion. We have then applied SIENA [2], an automatic technique for the estimation of atrophy rates, to different sets of atrophy simulated data with the presence of acquisition artefacts.

2 Method

The generation of ground-truth data from an MR scan for the evaluation of atrophy measurement techniques can be divided in two stages: simulation of atrophy and simulation of acquisition artefacts.

2.1 Simulation of Atrophy

The method for atrophy simulation can be divided into four phases: meshing of a labelled brain atlas; adaptation of the generated mesh to an individual MR scan; introduction of the individual-specific mesh into a FEM solver that will generate the simulated deformations; and the application of such deformations to the MR scan. More details about this method can be found in [5]. Figure 1 shows an example of atrophy simulation.

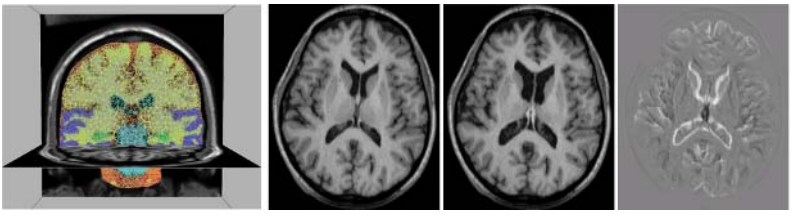


Fig. 1. From left to right: mesh, baseline, atrophy simulated scan and difference image

2.2 Simulation of Acquisition Artefacts

2.2.1 Rotational Motion

Rotational motion during a scan causes inconsistencies in the sampled k-space signal that result in blurring and ghosting in the image domain. The method we developed to simulate rotational motion is the following.

A motion-free 3D image volume (top left of Fig. 2) is converted into a complex image by treating the modulus image as the real part, and setting the imaginary component to zero. This complex image is then rotated by 5 degrees about an axis through the internal auditory meatus, simulating a slight nodding of the patient during the scan (bottom left of Fig. 2). Bicubic interpolation was used in the rotation, and voxels entering the volume were assigned the value zero. The original and rotated images were then both Fourier transformed into k-space (second column of Fig. 2). A motion-corrupted k-space was subsequently generated by assembling portions of the k-spaces corresponding to the un-rotated and rotated images (third column of Fig. 2). The motion-corrupted k-space was then transformed back to the image domain (final image of Fig. 2) to give the motion corrupted volume. The k-space samples were

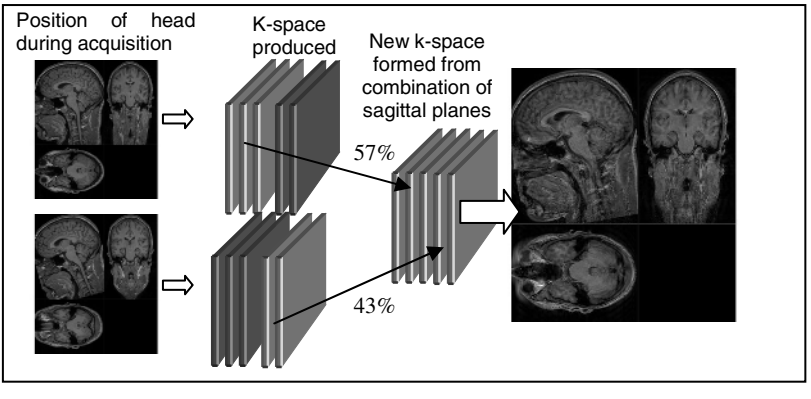


Fig. 2. Simulation of step rotational motion: motion is applied in the spatial domain to a motion-free image volume, and the k-spaces re-combined to generate a motion-corrupted image volume

combined as sagittal planes representing a sagittal MR acquisition orientation. Transverse or coronal planes could also have been simulated in these 3D volumes by appropriately assembling the final k-space.

The appearance of the motion artefacts is influenced by which region (its distance to the centre of the k-space) of k-space is being sampled as the motion takes place, and whether there is a single step motion, or a periodic motion. Our technique allows complete freedom in the selection of the time of the motion in the scan, and for the examples in this paper, we simulated step motion by combining the first 57% of the un-rotated k-space, with 43% of the rotated k-space. We simulated the motion just after half way through the acquisition as the artefact is more noticeable the nearer to the centre of k-space it occurs. Periodic motion, representing a repeated nodding of the head, was simulated by generating the motion-corrupted k-space from portions of the un-rotated and rotated k-space, alternating between the two at 44%, 48%, 52% and 56%. Figure 3 shows an example of the images resulting from simulated step and periodic rotational motion.

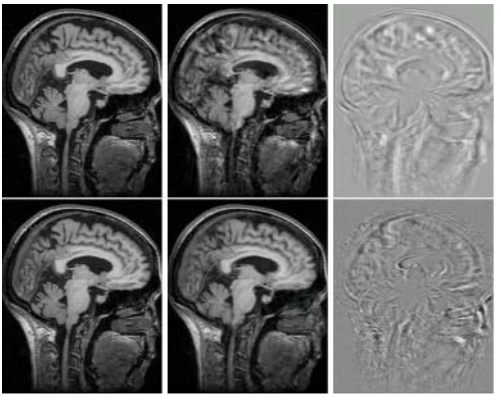


Fig. 3. Top: periodic motion, bottom: step motion. From left to right: Atrophy simulated image, atrophy plus motion artefact simulated images, difference image.

2.2.2 Pulsatile Motion

Another important artefact in MR images is pulsatile motion, which typically leads to streaks across the image in the slow-phase encoding direction. These artefacts arise from major blood vessels in the field of view, such as the carotid artery where it passes through the carotid siphon, and result from a signal in the vessel that varies from shot to shot during the acquisition. Simulated pulsatile motion artefacts were added to the simulated data as follows (see also Figure 4).

The 3D modulus images were converted to a complex image as described above, and four different coil views were obtained by multiplying the image by four separate coil sensitivity profiles. Because the pulsatility artefact is predominantly in the

slow- phase encode direction, each plane of k- space was treated as being acquired instantaneously (contrast being largely determined by the acquisition of the centre portion of k-space) and 1.73s apart from the next. A cylinder corresponding to the carotid artery was manually delineated from the image, and the intensity of this modulated according to a literature-based model of the flow in the carotid artery [11]. For each shot in the slow-phase encode direction, a different image was simulated by modulating the intensity of the carotid signal in the delineated cylinder. These images were transformed into the k-space, and one k-space plane from each image was combined to form a pulsatile-motion corrupted k-space based on an estimated heart rate of 60bpm. Because this is a one dimensional motion, it was implemented just for the lines containing the carotid, rather than the entire image, in order to reduce computational cost.

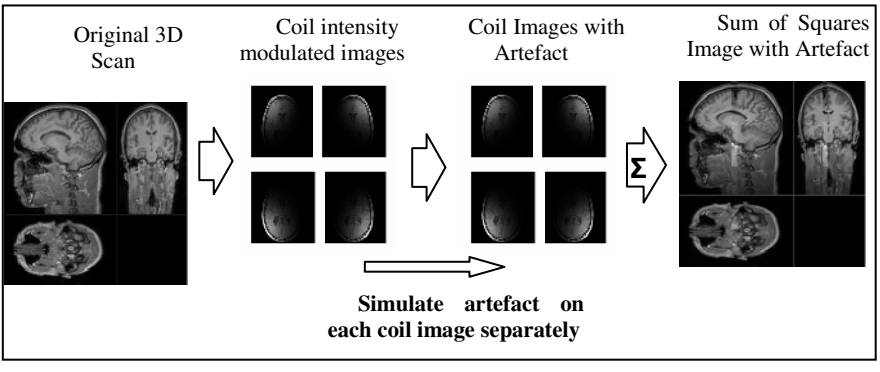


Fig. 4. Overview of simulation of flow artefact

The resulting image, as can be seen in Figure 5 has a simulated flow artefact running right to left. This is because we simulated the slow phase encode direction in the left-right direction as is common in 3D volumes used in dementia imaging. This is implemented in the simulation by using sagittal planes of k-space. However we could simulate any direction of acquisition which would alter the direction of the simulated artefact, and could also simulate more complex vessel geometries or flow properties.

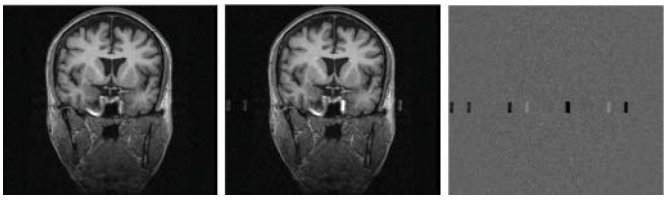


Fig. 5. From left to right: Atrophy simulated image; atrophy and flow artefact simulated image; difference image

3 Experiment

3.1 Design

We have designed an experiment in order to investigate the effect of the simulated acquisition artefacts in atrophy measurements provided by an automatic technique, SIENA (Structural Image Evaluation, using Normalization, of Atrophy), proposed by Smith et al. [2].

Firstly, this method segments brain from non-brain tissue in the head, and estimates the outer skull surface. Then, it uses these results to register the two brains with the skulls constrained. These images are then segmented, and the atrophy is estimated from the movement of the segmentation boundaries. The output is the estimation of the percentage of brain volume change between the two scans, as well as a flow image in which regions of local change are shown, as can be seen in Figure 5. The authors of this technique claim a median absolute error of 0.2% of the whole brain volume change in a validation study.

In this work, we have employed 7 MR scans of healthy elderly controls enrolled in a longitudinal research project at the anonymous Centre, anonymous Institute, anonymous University, anonymous country and underwent serial MRI, clinical and neuropsychological assessment over a period of at least one year. T1-weighted volumetric MR images were acquired on a 1.5 Tesla Signa Unit (General Electric, Milwaukee) using a 256*256 matrix to provide 124 contiguous 1.5mm coronal slices through the head (acquisition parameters: time to repeat, 15ms; time to echo, 5.4ms; flip angle 15°, field of view 24*24cm).

We have applied the atrophy simulation technique [6] to these 7 MR scans, generating two new images, *Atsim1* and *Atsim2*, for each subject scan (baseline) with different amounts of regional and global atrophy in several brain tissue compartments and structures, resulting in a dataset of 21 images. In this work, we only make use of the atrophy simulation's ground truth values for the whole brain volume change since SIENA only gives an estimation of the global atrophy rate. The mean and standard deviation of the simulated whole brain volume change percentages are the following: a reduction of 5.11% ± 0.23 for *Atsim1*; and a reduction of 10.84% ± 0.47 for *Atsim2*.

The next phase of the experiment applies the methods described in Section 2 to the dataset of atrophy simulated images in order to simulate the acquisition artefacts detailed above. Specifically, we simulate step and periodic rotational motion (see Section 2.1) for the 21 available images and pulsatile motion (see Section 2.2) for a subset of 7 images (three different cases and their corresponding atrophy simulated images), due to the time-consuming manual segmentation of the carotid artery needed to simulate this artefact. Furthermore, we have obtained, for a single case and its associated *Atsim1*and *Atsim2* images, 10 different step and periodic rotational motion simulated images varying the severity of the artefact (i.e. changing degrees of rotation).

Finally, we run SIENA on the whole dataset of simulated images, aiming at studying the effect of acquisition artefacts on atrophy measurements.

3.2 Results

Results are summarized in Table 1. We have computed the difference (diff1) between the whole brain atrophy measurement calculated by SIENA and the gold standard, for the images that contain both simulated atrophy and simulated artefacts. We have also computed the difference (diff2) in SIENA results obtained with just simulated atrophy, and with simulated atrophy plus simulated artefact. The main difference between both measures is that "diff1" accumulates SIENA errors in the estimation of atrophy as well as errors induced by the simulation of acquisition artefacts, while "diff2" is independent of the atrophy simulation stage.

Table 1. Effect of simulated acquisition artefacts. diff1: difference between SIENA results and the gold standard for atrophy and artefact simulated images. diff2: difference in SIENA results between just simulated atrophy and with simulated atrophy plus simulated artefact. diff1 and diff2 are percentages of whole brain volume change.

	No artefact	Periodic	Step	Flow
diff1	0.286±0.383	1.197±0.777	1.123±0.904	0.412±0.220
diff2	n/a	0.982±0.659	0.941±0.738	0.479±0.099

Both types of difference measures suggest that the presence of artefacts in the images has a substantial impact on atrophy measurements obtained with SIENA. We can observe in Table 1 that rotational motion artefacts, either step or periodic, degrade the performance of SIENA more than flow artefacts. This is due to the limited region where pulsatile artefacts are present, mostly the temporal lobes, while SIENA only provides a global estimation of the whole brain volume change. It must be pointed out that the accuracy obtained with SIENA in atrophy measurements without artefacts (first column of Table 1) are consistent with results provided by the authors of this technique (accuracy of 0.2% of whole brain volume change).

Figure 6 shows the impact on the error in the SIENA atrophy measurement as a function of magnitude of the rotational motion (stop motion from 0° to 10° degrees in intervals of 1 degree). There is a strong correlation between the increase in atrophy measurement errors and the severity of the simulated artefact. The expected difference in whole brain volume change between AD and normal aging over a 12 month period is around 1.5%. Results in Figure 6 suggest that rotational motions larger than 5 degrees of rotation could induce artefacts around the same order of these expected differences.

4 Conclusions

We have presented the techniques we have developed to simulate the two most common acquisition artefacts in MR scans, rotational (both step and periodic) and pulsatile motion, in order to make ground truth data for atrophy measurement techniques more realistic. The methodology can be used to simulate many different

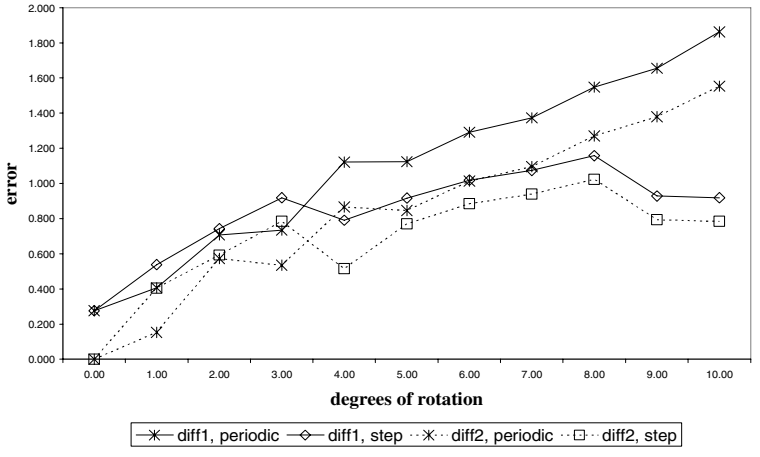


Fig. 6. Effect on SIENA results varying the severity of the rotational simulated artefacts

magnitudes of artefacts, and can be applied to many different atrophy quantification methods.

Only one atrophy quantification method has been investigated. The impact is likely to be different with different algorithms, and especially for local atrophy measures (e.g.: from a non-rigid registration algorithm) or cortical thickness measurement methods. The value of the approach is in predicting the sensitivity of methods to motion artefact (potentially valuable in a power analysis used in the design of a clinical trial), and also in comparing the relative merits of alternative atrophy measuring algorithms in the presence of common image artefacts.

References

1. Freeborough, P., Fox, N.: The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI. IEEE Transactions on Medical Imaging 16 (1997) 623–629
2. Smith, S., Stefano, N.D., Jenkinson, M., Matthews, P.: Normalized accurate measurement of longitudinal brain change. Journal of Computer Assisted Tomography 25 (2001) 466–475
3. Davatzikos, C., Genc, A., Xu, D., Resnick, S.: Voxel-based Morphometry using the RAVENS maps: methods and validation using simulated longitudinal atrophy. Neuroimage 14 (2001) 1361–1369
4. Freeborough, P., Fox, N., Kitney, R.: Modeling brain deformations in Alzheimer's disease by fluid registration of serial 3D MR Images. Journal of Computer Assisted Tomography 22 (1998) 838–843
5. Camara, O., Schweiger, M., Scahill, R.I., Crum, W.R., Sneller, B.I., Schnabel, J.A, Ridgway, G.R., Cash, D.M., Hill, D.L.G., Fox N.C.: Phenomenological model of diffuse global and regional atrophy using Finite-Element methods. IEEE Transactions on Medical Imaging (2006). In press.

6. Howarth, C., Hutton, C., Deichmann, R.: Improvement of the image quality of T1-weighted anatomical brain scans. Neuroimage 29 (2005) 930–937
7. Atkinson, D., Hill, D.L., Stoyle, P.N., Summers, P.E., Clare, S., Bowtell, R., Keevil, S.F.: Automatic compensation of motion artifacts in MRI. Magnetic Resonance Medicine 41 (1999) 163–170
8. Manduca, A., McGee, K.P., Welch, E.B., Felmlee, J.P., Grimm, R.C., Ehman, R.L.: Autocorrection in MR imaging: adaptive motion correction without navigator echoes. Radiology 215 (2000) 904–909
9. Blumenthal, J.D., Zijdenbos, Al, Molloy, E., Giedd, J.N.: Motion artifact in magnetic resonance imaging: implications for automated analysis. Neuroimage 16 (2002) 89–92
10. Preboske, G.M., Gunter, J.L., Ward, C.P., Jack Jr., C.R.: Common MRI acquisition non-idealities significantly impact the output of the boundary shift integral method of measuring brain atrophy on serial MRI. Neuroimage (2006) In press.
11. Holdsworth, D., Norley, C., Frayne, R., Steinman, D., Rutt, B.: Characterization of common carotid artery blood-flow waveforms in normal human subjects. Physiological Measurement 20 (1999) 219–240

Non-rigid 2D-3D Registration with Catheter Tip EM Tracking for Patient Specific Bronchoscope Simulation

Fani Deligianni, Adrian J. Chung, and Guang-Zhong Yang

Department of Computing, Imperial College London
{fani.deligianni, ajchung, g.z.yang}@imperial.ac.uk
http://www.doc.ic.ac.uk/~gzy/vis-a-ve

Abstract. This paper investigates the use of Active Shape Models (ASM) to capture the variability of the intra-thoracic airway tree. The method significantly reduces the dimensionality of the non-rigid 2D/3D registration problem and leads to a rapid and robust registration framework. In this study, EM tracking data has been also incorporated through a probabilistic framework for providing a statistically optimal pose given both the EM and the image-based registration measurements. Comprehensive phantom experiments have been conducted to assess the key numerical factors involved in using catheter tip EM tracking for deformable 2D/3D registration.

1 Introduction

The clinical practice of minimal invasive surgery has advanced rapidly in recent years due to improved therapeutic outcome coupled with reduced patient trauma and hospitalization. The technique, however, requires extensive manual dexterity and hand-eye coordination due to the complexity of instrument control and a loss of 3D vision and tactile feedback. In bronchoscope, for example, these skills are typically acquired based on the apprentice scheme or inanimate plastic models. The use of high-fidelity simulators provides an attractive alternative which offers a more complete training curriculum by incorporating usual, as well as rare pathological cases at a significantly lower cost and without the inference of routine patient care.

The prerequisite of constructing high-fidelity simulation models is the provision of visually realistic rendering based on patient-specific biomechanical models to recreate an environment that is close to real surgical scenes. This approach of patient specific simulation has attracted considerable interests in recent years [1, 2], but photorealism and biomechanical fidelity remain active research topics [3-5]. In bronchoscope simulation, an essential requirement for photorealistic rendering and modeling is accurate 2D/3D registration of the 2D bronchoscopy video with 3D tomographic data. Despite the success of rigid 2D/3D registration achieved in recent years, existing research has shown that the accuracy of the current registration techniques is compromised due to the deformation of the airways [6]. During respiration, the intra-thoracic tree is affected by both respiratory and cardiac motion and it undergoes significant anatomical changes even during normal breathing. This explains why most

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 281 – 288, 2006.

of the current techniques become problematic in registering distal airways where structural deformation is at its greatest.

The incorporation of airway deformation to the existing 2D/3D registration framework, however, is by no means trivial. This is due to the potentially large search space involved in simultaneous tracking of the camera pose and morphological deformations. Issues concerning mesh folding and the sensitivity of the similarity function to lighting and local minima represent a major challenge. With the recent advent of catheter tip EM trackers, research is now directed towards using the device during the bronchoscope procedure to simplify the 2D/3D registration process [7]. These devices typically have 5 degrees-of-freedom (DoF) and have also been used in a number of image-guided applications such as for precision biopsy [8] and spine surgery [9].

As a first glance, it may seem that the use of 5 DoF catheter tip EM tracker reduces the original 2D/3D registration to only a 1D problem. It has been postulated that in the presence of airway deformation, one can readily use the EM tracking result to recover the unknown axial camera rotation [10]**.** After careful *in vivo* experiment, however, the true extent of the problem has been recognized and it has been found that the effect of respiratory motion in previous studies has not been fully appreciated. This is because the EM tracker reports the position and orientation of the bronchoscope camera in relation to the fixed, world coordinates. When this information is used to guide the registration algorithm to the static airway acquired from tomographic data, a large displacement due to airway deformation can be observed. When this EM tracking information is used to initialize the 2D/3D registration, it has often been found that the initial view can be completely outside of the bronchial tree. This initialization is in fact inferior to the traditional image based approach, as in this case the camera view is always in synchrony with the moving frame of reference of the deforming airways. For this reason, direct mapping of the EM tracking measurements to the static pre-operative CT images is not be appropriate, and its practical use must be integrated with the deformation of the airways.

The purpose of this paper is to examine the use of Active Shape Model (ASM) to capture the variability of the intra-thoracic tree. In this way, it is expected that the dimensionality of the non-rigid 2D-3D registration problem can be significantly reduced, thus leading to a rapid and robust registration framework. Since it is difficult to obtain the ground truth for *in vivo* data, this study is focused on the use of comprehensive phantom validation to assess the key numerical factors involved in using catheter tip EM tracking for deformable 2D/3D registration.

2 Methods

2.1 Modeling the Respiratory Motion of the Intra-thoracic Tree

Physically-based models are controlled by the laws of motion dynamics and corresponding tissue properties. They generally result in accurate bio-mechanical representation of the deformation. However, the practical implementation of this approach can be difficult due to the lack of detailed tissue mechanical properties, as

well as the prohibitive computational cost involved. One alternative is to use general shape modeling such as the NURBS based surface representation for the intra-thoracic tree [5]. In addition to the issue of biomechanical fidelity, the main disadvantage of this approach is the mathematical complexity it introduces and the large number of unrestricted parameters that control the shape of the NURBS.

In this study, the ASM is used to provide subject specific modeling of the deformation of the airways during respiration. ASM provides an effective means of capturing statistically plausible variations of an object shape through a number of observations [11]. ASM is based on Point Distribution Models (PDMs), where point correspondence between a set of registered images is established and the variability of the shape can be expressed through a set of a linear combination of the mean shape $\bar{s}$ derived from the training set and its modes of variation. That is,

$$s = \bar{s} + Pb \tag{1}$$

where P is the eigenvectors and b the parameter that varies to create new shapes that are restricted by the main modes of variation. With ASM, simplification of the model is possible by selecting only the principal modes of variation. The first mode of variation, for example, typically describes the main motion of the tracheobronchial tree due to respiration, whereas other modes are related to high-order local variabilities.

It is important to note that the success of ASM depends on the number and quality of correspondent landmarks across the training shapes. To establish correspondence between a set of control points across the 3D meshes, a semi-automatic method was used. The skeleton of the airway tree was first extracted from the 3D CT data. The bifurcations of the airway tree were then used as the landmarks for establishing shape correspondences. For each of the skeleton points, the surface perpendicular to the skeleton axis was defined and its intersection with the 3D mesh was estimated. The application of ASM to the skeletal and surface control points results in intermediate, deformed airway structures. To generate a smooth 3D mesh of the airways, Radial Basis Functions (RBFs) are used. RBFs provide an easily controllable behaviour based on a sparse and irregular set of control points and thus they are suitable for deforming the tracheo-bronchial mesh.

2.2 Incorporating Temporal Correspondence

The use of ASM allows the incorporation of deformable modeling into the registration process. To further improve the registration accuracy, temporal tracking based on the Condensation algorithm has been used. Particularly, the condensation algorithm permits the use of multi-modal probability distributions that is essential to track the camera position under airway deformation and image artifacts [12]. Mathematically, the condensation algorithm does not explicitly represent the density function. It uses instead statistical factor sampling based on the Bayesian theory. During statistical factored sampling, a sample set is generated from the prior density. A weight is subsequently assigned to each particle according to the observation density. The weighted point set is then used as a representation of the *a posteriori* density. Subsequently, by evaluating the given moments of the state density, it is possible to estimate the most probable state. During prediction, the method involves the evolution of the state vector with time. Sampling methods based on spatial

Markov processes are generally used for the Condensation method to represent the dynamics of a moving camera.

For the purpose of this study, the observation/measurement model needs to combine both the EM tracker measurements and the image-based registration results. Generally, the observation density $p(z_t \mid x_t)$ defines the likelihood that a state x_t causes the measurement z_t. This is typically achieved by assuming that the measurements are statistically independent both mutually and with respect to the process model. Furthermore, both the EM tracker and the image-based registration measurements are imprecise. In this study, the uncertainty is quantified with a Gaussian probability density function $p \sim \exp(0, \sigma)$, where the variance σ, depends on the reliability of the different sensors measurements. The observation density for a particular state is estimated by integrating over the space of sensors measurements:

$$p(z_t \mid x_t) = \sum_m p_{z_t^m} \tag{2}$$

For 2D/3D registration, a *pq*-space based 2D/3D registration technique developed by Deligianni *et al* has been used [1]. The state of the system has been defined as the 6DoF pose of the camera. However, since EM tracking data are available only for 5 out of 6 DoF, the last degree of freedom is tracked independently.

Finally, for the process model, a second order auto-regressive model was used to describe the motion of the endoscopic camera [13]. The auto-regressive model takes into account that during bronchoscope navigation, motion occurs within a bounded area, and a rapidly moving camera is expected to slow down or change in direction rather than accelerate further . A second-order auto-regressive model is described by the following equation:

$$x_t = A_1 x_{t-1} + A_2 x_{t-2} + Bwt \tag{3}$$

where A_k represents the collection of damped harmonic oscillators associated with vibrational modes and w the white noise with covariance coefficient B. This ensures smooth navigation of the camera. With modular training, multiple training sets can be used to obtain a more general model of the motion behavior .

Modular training facilitates the incorporation of multiple sequences from different subjects as well as the incorporation of the dynamic behavior of camera navigation in different parts of the tracheo-bronchial tree [13, 14]. Based on the principles of the maximum likelihood learning, the parameters of the second order autoregressive model can be estimated . The motion characteristics such as the coefficients A_k, the mean value $\overline{X}$ of the state and the random component B are obtained from a number of different training sequences derived directly from the 5 DoF EM data, while for the sixth degree of freedom manually collected data has been used.

2.3 Validation

To validate the proposed technique, a deformable airway phantom made of silicon rubber and painted with acrylics was used. The inner surface was coated with silicon-rubber mixed with acrylic to give it a textured specular finish that is similar to

the bronchial lumen. The phantom surface with finished with red color splashes to imitate the blood and texture discrepancies during a bronchoscope examination. The distal ends of the airway-phantom have been fixed and a balloon was located below the main bifurcation to simulate respiratory induced the airway deformation. The extent of the airway motion is regulated by a valve to control the amount of air in the balloon. The tomographic model of the phantom was scanned with a Siemens Somaton Volume Zoom 4-channel multi-detector CT scanner with a slice thickness of 1 mm and in-plane resolution of 1 mm. A total of six *CT* volumes were acquired at different stages of the motion cycle. An Olympus bronchoscope (Olympus BF Type, with a field of view 120^0) operating in PAL recording mode (25fps) was used to capture the 2D bronchoscope video. Prior to 2D/3D registration, pre-processing of the videos was necessary in order to alleviate the effects of interlacing, lens distortion and image artefacts [15].

To track the pose of the bronchoscope camera, an NDI (Northern Digital Inc, Ontario) Aurora EM tracker is used. The Aurora system is particularly suited to tracking flexible endoscopes during minimally invasive intervention. This typically employs a single coil sensor due to size restrictions imposed by the dimensions of the biopsy channel or catheter. Having only a single coil, the sensor is limited to reporting position and direction only (*i.e.* 5 DoF with unknown axial rotation). The working volume of the Aurora system is a 500mm cube located 50mm from the field generator. In an environment that is free from electromagnetic interference, the Aurora has a static positional accuracy of 0.9-1.7mm and an orientation accuracy of 0.5 degree .

In order to provide the ground truth data for airway deformation, a 6 DoF EM tracker was used to track the deformation of the balloon so that the deformation of the phantom airway can be indexed back to the 3D CT data. Based on the assumption that these points are moving in parallel for each positional value of the 6 DoF EM tracker the phase of the respiratory motion can be estimated as:

$$b_1 = \vec{M} \cdot p_r \cdot \left(p_r \cdot p_r^T \right)^{-1} \tag{4}$$

where p_r is the eigenvector that corresponds to the point s_r of the mean shape that is nearest to the position of the 6DoF EM tracker and $\vec{M}$ is the vector between these points.

3 Results

In Fig. 1 the recovered deformation from the ASM model is presented. During 2D/3D registration, this has been used to directly deform the tracheo-bronchial mesh by tuning the first principal component of the ASM. The markers point out the location of the frames depicted in Fig. 2, which provides a qualitative validation of the recovered coefficients of the ASM. In this figure, column (a) represents the original bronchoscope video frames and column (b) illustrates the rendering pose by directly using the EM data. It is evident that the position of the camera changes rapidly due to the respiratory motion in the world coordinates, and a direct use of the EM data can result in initial camera position located completely outside of the bronchial tree. The result of incorporating ASM to the 2D/3D registration framework is shown in Fig. 2(c), which defines the position and orientation of the camera such that the

remaining axial rotation can be recovered as shown in Fig. 2(d) . Finally, Fig. 2(e) demonstrates the results by using Fig. 2(d) to initialize the 6 DoF *pq*-based registration [1].

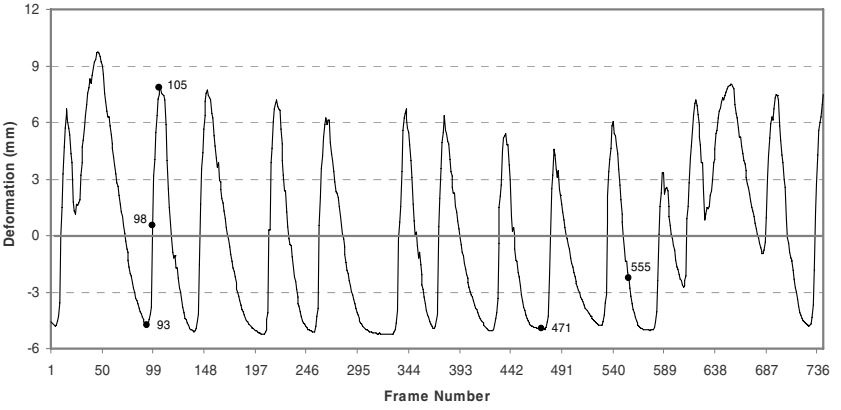


Fig. 1. Recovered deformation that can be used to directly deform the tracheo-bronchial mesh by tuning the first principal component of the ASM

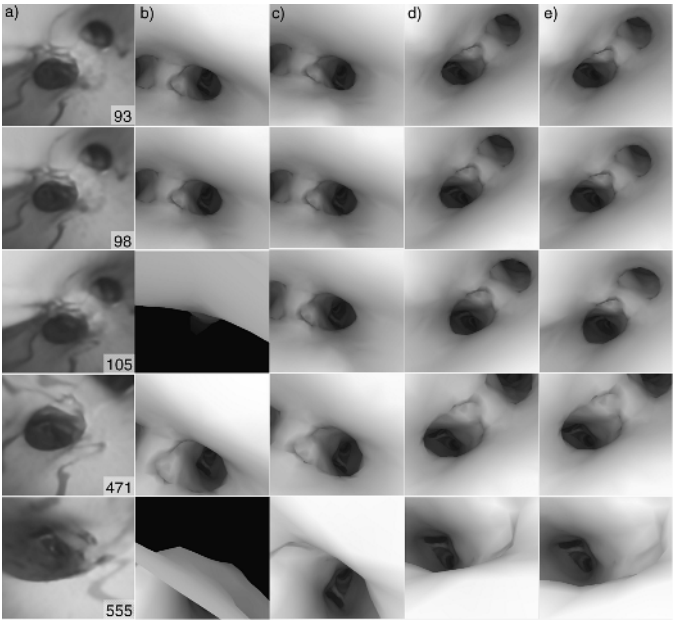


Fig. 2. Example bronchoscope video frames (a), and the corresponding registration results by a direct application of the EM tracking information (b), and by deformation modeling with ASM (c). Columns (d)-(e) illustrates the registration results after axial rotation recovery and subsequent 6 DoF *pq*-based registration, respectively

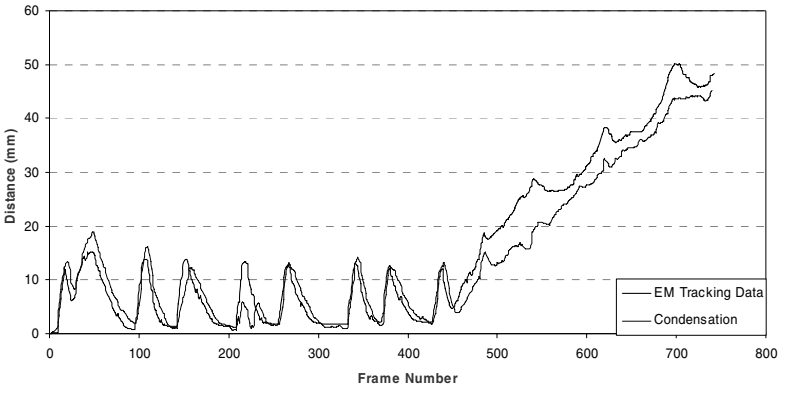


Fig. 3. Euclidean distance between the first and subsequent camera position as it has been predicted from the Condensation algorithm with relation to the EM tracking data

To provide a quantitative measure of the tracking algorithm described above, Fig. 3 illustrates the results by applying the Condensation algorithm the entire video to segment as compared to the ground truth data recovered from the EM data. Temporal tracking incorporates both the measurements of the EM tracker andc the image-based registration results, and thus it is able to recover when either of them has failed.

4 Discussion and Conclusions

Despite the prevalence of tissue deformation in most surgical navigation and planning applications, its incorporation into the 2D/3D registration framework has been limited. This is due to the large degrees of freedom involved in the registration parameters, which make a stable implementation of the algorithm difficult. In this paper, we have developed a non-rigid 2D/3D registration framework that models the respiratory motion of the intra-thoracic tree based on ASM. ASM captures the intrinsic variability of the airways across different phases of the respiratory motion and it constrains the motion they represent to be specific to the subject studied. This allows the subsequent non-rigid registration implemented in a much reduced dimensional space, and thus greatly simplifies the 2D/3D registration procedure that incorporates temporal correspondence and local adjustment. The detailed phantom validation of the method in this study demonstrates the potential clinical value of the technique.

References

1. Deligianni F, Chung A, Yang GZ. Patient-Specific Bronchoscope Simulation with pq-Space-Based 2D/3D Registration. *Computer Aided Surgery* 2004; 9(5): 215-226.
2. Chung AJ, Deligianni F, Shah P, Wells A, Yang GZ. Vis-a-Ve: Visual Augmentation for Virtual Environments in Surgical Training. in Proceedings of Eurographics / IEEE VGTC Symposium on Visualization, Leeds, UK, 2005, 101-108.

3. Chung AJ, Deligianni F, Shah P, Wells A, Yang GZ. Patient Specific Bronchoscopy Visualisation through BRDF Estimation and Disocclusion Correction. *IEEE Transactions of Medial Imaging* (accepted for publication), 2005.
4. Deligianni F, Chung A, Yang GZ. Non-Rigid 2D/3D Registration for Patient Specific Bronchoscopy Simulation with Statistical Shape Modelling. *IEEE Transactions on Medical Imaging* (under review), 2005.
5. Garrity JM, Segars WP, Knisley SB, Tsui BMW. Development of a Dynamic Model for the Lung Lobes and Airway Tree in the Ncat Phantom. *IEEE Transactions on Nuclear Science* 2003; 50(3): 378-383.
6. Mori K, Deguchi D, Sugiyama J, Suenaga Y, Toriwaki J-i, Jr. CRM, 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* 2002; 6(3): 181-336.
7. Hautmann H, Schneider A, Pinkau T, Peltz F, Feussner H. Electromagnetic Catheter Navigation During Bronchoscopy: Validation of a Novel Method by Conventional Fluoroscopy. *Chest* 2005; 128(1): 382-387.
8. Banovac F, Glossop N, Lindisch D, Tanaka D, Levy E, Cleary K. Liver Tumor Biopsy in a Respiring Phantom with the Assistance of a Novel Electromagnetic Navigation Device. in Proceedings of Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI02), Tokyo, Japan, 2002, 200-207.
9. Sagi HC, Manos R, Benz R, Ordway NR, Connolly PJ. Electromagnetic Field-Based Image-Guided Spine Surgery Part One: Results of a Cadaveric Study Evaluating Lumbar Pedicle Screw Placement. *Spine* 2003; 28(17): 2013-2018.
10. Mori K, Deguchi D, Akiyama K, Kitasaka T, Maurer CR, Suenaga Y, Takabatake H, Mori M, Natori H. Hybrid Bronchoscope Tracking Using a Magnetic Tracking Sensor and Image Registration. in Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI05), Palm Springs, California, USA, 2005, 543-550.
11. Cootes TF, Taylor CJ, Cooper DH, Graham J. Active Shape Models: Their Training and Application. *Computer Vision and Image Understanding* 1995; 61(1): 38-59.
12. Isard M, Blake A. Condensation-Conditional Density Propagation for Visual Tracking. *International Journal of Computer Vision* 1998; 29(1): 5-28.
13. Deligianni F, Chung A, Yang GZ. Predictive Camera Tracking for Bronchoscope Simulation with Condensation. in Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI05), Palm Springs, California, USA, 2005, 910-916.
14. North B, Blake A, Isard M, Rittscher J. Learning and Classification of Complex Dynamics. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2000; 22(9): 1016-1034.
15. Yang GZ, Burger P, Firmin DN, Underwood SR. Structure Adaptive Anisotropic Image Filtering. *Image and Vision Computing* 1994; 14(2): 135-145.

Anatomical Modelling of the Musculoskeletal System from MRI

Benjamin Gilles, Laurent Moccozet, and Nadia Magnenat-Thalmann

MIRALab, University of Geneva, CH-1211 Geneva, Switzerland,
{gilles, moccozet, thalmann}@miralab.unige.ch

Abstract. This paper presents a novel approach for multi-organ (musculoskeletal system) automatic registration and segmentation from clinical MRI datasets, based on discrete deformable models (simplex meshes). We reduce the computational complexity using multi-resolution forces, multi-resolution hierarchical collision handling and large simulation time steps (implicit integration scheme), allowing real-time user control and cost-efficient segmentation. Radial forces and topological constraints (attachments) are applied to regularize the segmentation process. Based on a medial axis constrained approximation, we efficiently characterize shapes and deformations. We validate our methods for the hip joint and the thigh (20 muscles, 4 bones) on 4 datasets: average error=1.5mm, computation time=15min.

1 Introduction

For the diagnosis, the surgical planning and the post-operative assessment of musculoskeletal disorders, the automatic segmentation of the patient musculoskeletal system is important for orthopaedists, biomechanicians and kinesiologists that would like to simulate, visualize and navigate through articulations with a minimum amount of manual tasks. As stressed by [1], [2] and [3], usual simplified models (stick-figures, muscle action lines) are not able to take into account large attachment areas, as well as global constraints such as volume preservation and non-penetration; although they are important biomechanical parameters. Therefore, the relationship between musculoskeletal dynamics and organ shapes need to be better studied through image segmentation. Magnetic Resonance Imaging (MRI) is a flexible modality for imaging both soft and bony tissues non-invasively. However, due to the large amount of textural information, noise, low-resolution, organ imbrication, and large spatial variability; automatic, fast and robust musculoskeletal segmentation is a difficult task. As a consequence, existing methods for musculoskeletal modelling are interactive [2] [4] [1] [3] and therefore time-consuming. A common approach to constrain a segmentation process is to use prior information: shape constraints rely on assumptions about surface regularity (smoothness, curvature) and variability across the population, while topological constraints exploit prior knowledge about organ interrelationships. Contrary to traditional pixel-based segmentation (e.g. level-sets) and registration (e.g. FFD) methods, these constraints can be efficiently applied

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 289–296, 2006.

on deformable models [5]. Physical-based simulation approaches (volumetric), including finite element or finite volume are appropriate for enforcing mechanical constraints, but their computational cost would exclude time-efficient and user-controlled segmentation. In this context, we propose to use scalable discrete deformable surfaces (simplex meshes), that can benefit from efficient geometric methods, popular in computer graphics applications.

Simplex meshes were first described by H. Delingette [6] for constrained 3D shape reconstruction and segmentation, and extended in 4D by J. Montagnat [7], with application to the heart (single model and resolution). A k-simplex mesh is defined by a set of vertices and a connectivity function (each vertex is connected to exactly k+1 neighbors). In this paper, we use 2-simplex meshes (dual to triangle meshes). Mesh geometric quality (uniformity of vertices repartition) and topological quality (uniformity of edge number among faces) are improved using simple topological operators. The most interesting property is its simple geometric description: three parameters (two barycentric coordinates plus the curvature) uniquely define vertex positions from their three neighbors. Based, on these parameters, smoothing and shape constraint forces are computed to regularize the segmentation. External forces are obtained through 1D registration of intensity profiles (at vertex positions and in normal direction) with generic profiles from a reference segmentation. The external force field is regularized using a local smoothing and global regularization based on the closest affine transformation.

Discrete models are commonly considered as punctual masses evolving under the Newtonian law of motion. The Newton equation leads to a first-order differential equation system relating forces to particle state (velocity and position). After forces evaluation and time discretisation, particle state can be explicitly resolved (forward Euler, Runge-Kutta) with tight time step restrictions for ensuring stability. We prefer the more stable implicit (or backward) scheme [8] [9], that however requires the resolution of a large sparse linear system. To simplify force derivatives evaluation, we consider that forces have a independent anisotropic action on each particle. Non-penetration constraints are based on collision handling techniques [10]. In this paper, we apply a hierarchical collision detection scheme based on 18-discrete oriented polytope(DOP) quadtrees [10]. We perform collision correction and response on particle positions, speeds and accelerations, such as in [9].

From these background studies, we propose improvements in terms of computational speed and robustness by extending the simplex mesh framework with a multi-resolution scheme, topological constraints and medial surfaces. We show that the medial axis is suited for muscle shape analysis.

2 Methods

2.1 Multi-resolution Scheme

The use of levels of details (LODs) aims at reducing system complexity and sensitiveness to local solutions. As shape constraints spatial influence depends

on the resolution level, the idea is to quickly propagate forces from lower resolutions to a current simulation LOD in order to get multiscale regularization. In addition, collision detection (often considered as the bottle-neck for simulation) is more efficiently performed on coarse LODs, collision response being passed to fine LODs. Indeed, exact contact computation is most of the time not relevant as fat separates organs. The tessellation of dual triangle meshes leads to a systematic and computationally efficient LOD generation scheme for simplex meshes (linear combination of vertex positions). Shape features are preserved as low level vertices are contained in higher levels. The number of vertices is quadrupled when increasing the resolution. During the simulation, forces from lower resolutions are linearly combined like vertex positions and added to current resolution forces. Even if this is not exact (vertices having moved relatively to their neighbors, since resolution increase), the estimation is still relevant, assuming that shape constraints have enforced mesh local regularity.

2.2 Topological Constraints

The human musculoskeletal anatomy exhibits various organs interrelationships: muscles are attached to bones, they can merge into common tendon units, fascia binds muscles and enforces frictionless contact between them. We have developed a spline-based method for generating attachment areas, so that the placement and adjustment of areas have a reduced number of degrees of freedom. Spline control points are projected onto bone surfaces, while soft-tissues vertices are attached to the spline through curvilinear coordinates. These vertices are constrained using mass modification [8]: $\boldsymbol{M^{-1}} = \mathbf{0}$. For individualization, we wrap splines from a generic model using spline control point barycentric coordinates on bone surface. As shown in [11], this approach is valid for most attachments as they rely on bone geometrical features. For merging deformable models, we sum forces and masses of the attached vertices. This is interesting for modelling common tendons (e.g. quadriceps femoris tendon), but also specific parts with high curvature, where smoothing forces are not appropriate (e.g. attachment between the adductor magnus and its inferior tendon). Fascia is modelled through collision handling: once generic collisions and proximities have been detected, they are subsequently used as a reference (springs) to smoothly enforce the relative position between models. In other words, we deform generic contacts.

2.3 Radial Constraints

Muscles generally have a smooth and tubular shape. Hence, they can be efficiently represented by an underlying piece-wise action line [1] [4], where isotonic contractions are modelled through action line shrinking/stretching, and isometric contractions through radial constraints applied to a wrapped surface. This has been applied in biomechanics for calculating joint moment arms [3]. However, muscles with large attachment areas and/or several origins/insertions require many action lines. We propose a continuous representation using medial surfaces, leading to a continuous radial regularization of the overlying surface

and enhanced mechanisms for detecting collisions and characterizing shapes. Anchored medial surfaces (M-reps) have been successfully used in [5] [12] for constraining deformable model-based segmentation. The medial axis transform (computation of maximal ball centers and radii inside an object [13]) is an invertible transform that allows an efficient reduction of parameters dimension. Several methods have been presented for approximating the medial axis (MA), based on Voronoi diagrams, on distance maps, or on thinning.

We propose an iterative method based on forces applied to constrained deformable surfaces. Hook's spring interaction forces $\boldsymbol{f_j^*}$ are applied to medial surface vertices for MA approximation (j indexes medial vertices). Reciprocally, forces $\boldsymbol{f_i}$ are applied to overlying surface vertices for MA-based shape regularization (i indexes model vertices). Each model vertex $\boldsymbol{P_i}$ is associated to a medial surface point $\boldsymbol{P_{i\perp}}$ with a certain radius r_i. Radii are linearly interpolated between medial surface vertices $\boldsymbol{Q_j}$ of radius R_j to allow continuous force definition (leading to a smooth reconstructed surface from the MA). Given the weights w_{ij}, we have $\boldsymbol{P_{i\perp}} = \sum_j w_{ij}.\boldsymbol{Q_j}$ and $r_i = \sum_j w_{ij}.R_j$. From a reference state (Fig. 1a), weights are obtained by projecting orthogonally $\boldsymbol{P_i}$ onto the MA, and radii by the weighted mean $R_j = \sum_i w_{ij}.P_iP_{i\perp} / \sum_i w_{ij}$. Weights w_{ij} and radii R_j are subsequently memorized to compute radial forces for deformed states (Fig. 1b). Model forces $\boldsymbol{f_i}$ are simply defined by $\boldsymbol{f_i} = k.(1 - r_i/P_iP_{i\perp})\boldsymbol{P_iP_{i\perp}}$ where k is the spring stiffness (or force weight). By applying the momentum conservation law [9], we calculate the force contribution $\boldsymbol{f_{ij}}$ for one spring i at $\boldsymbol{Q_j}$ such as $\boldsymbol{f_{ij}} = -w_{ij}.\boldsymbol{f_i} / \sum_j w_{ij}^2$. We average all spring forces associated to a MA vertex: $\boldsymbol{f_j^*} = mean_{\,i}(\boldsymbol{f_{ij}})$. To obtain the reference medial surface for a given model, we first construct and attach a simplex plane, based on muscle attachments(Fig. 3d). After cropping, we simulate the medial surface with radial and smoothing forces, while projecting model vertices and updating radii at each timestep. The medial surface quickly converges to the medial axis. We measure the error between the reconstructed surface from the medial axis and the desired initial model shape by calculating the mean of $|r_i - P_iP_{i\perp}|$. The average error for all tested models is 0.6mm. This value remains stable during the segmentation when applying radial forces to both models and axis.

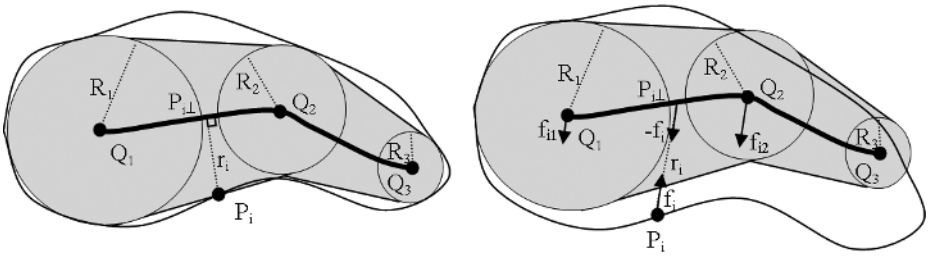


Fig. 1. Surface model and its associated medial axis in: A) A reference state; B) A deformed state; In grey, the reconstructed surface from the medial axis

In addition, we use medial axis for collision detection: 18-DOP quadtrees hierarchies, inflated according to radii, are generated/updated for medial axis cells. Muscles vertices at a certain resolution are subsequently tested towards DOPs and projected to medial axis for the inclusion test (comparison with interpolated radii r_i). Independently of the model resolution, medial axis based-collision detection is around 3 times faster than surface-based collision detection. On top of MA based-collision detection, multi-resolution collision handling (section **2.1**) is used to improve computation time (by a factor of 7 per LOD). Given a current resolution, we use its first coarser LOD for collision handling in order to minimize errors.

3 Results

3.1 Automatic Segmentation of the Musculoskeletal System

MRI protocol definition has been done in close collaboration with physicians from radiology and orthopaedics. The goal was to obtain images carrying sufficient information with clinically achievable protocols (fast). The final protocol was T1-weighted spin echo with TR/TE=578/18ms, Matrix/ FOV=512x512/40cm and slice thickness=2mm to 10mm between the foot the the iliac crests (five series in total, ∼150 slices, ∼30 minutes).

Based on interactive segmentation, we have reconstructed a generic model of the hip and the thigh, composed of a skin, 20 muscles, 4 bones and the corresponding attachment splines. After topological optimization, the different LODs (3 for soft-tissues, 4 for bones) and medial surfaces were generated. The final surface model (highest resolution) is composed of 71328 vertices for muscles and 85100 vertices for bones. Alternatively, muscles can be represented by their medial surfaces (1 resolution, 3821 vertices) with an average error of 0.6mm. The compression factor (parameter dimension reduction) is equal to (3x71328)/(4x3821)=14 (3 dimensions for model points, 4 for axis points, including radii).

Before automatic individualization, the generic model is coarsely initialized using ten manually placed landmarks corresponding to anatomical landmarks, and using thin-plate-spline (TPS) interpolation. After this step, bones are deformed (Euler implicit integration) from the coarse level to the fine level using multi-resolution internal forces (shape and smoothing constraints), non-penetration constraints (multi-resolution collision handling) and intensity profile-based external forces. Muscles are subsequently individualized (Fig. 2) as follows: A) attachment splines are initialized on bone surfaces from their generic barycentric coordinates; B) a skeleton-driven deformation algorithm [14] (skinning) is applied to generic muscles and medial surfaces according to joint angles; C) soft-tissues/medial surfaces are deformed using internal forces (radial/shape and smoothing constraints), proximity constraints (deforming contacts), and skin surface matching (gradient-based external forces); D) soft-tissues are deformed from the coarse level to the fine level using internal forces, proximity/non-penetration constraints and intensity profile-based external forces. The best

intensity profile size and resolution have been experimentally defined from interactively segmented models by minimizing excursions. Being more variable, the external part of the profile is less relevant (thus shorter). The intensity profile search depth decreases from 20mm to 5mm during segmentation (Fig. 2). We found that normalized cross-correlation is the most robust metric for intensity profile similarity measure. The affine regularization contribution decreases from 100% to 20%. For fine resolutions, internal forces from coarse levels are propagated (global/local constraints) as presented in **2.1**. Muscle shape constraints are derived from their medial surfaces with generic radii and weights. However, radii are updated for fine levels to give more freedom to surfaces. Radial forces are applied simultaneously to surfaces and to medial axis with an equal contribution. For coarse levels, generic proximity constraints are applied (deforming contacts). These constraints are released for the finest level, where multi-resolution medial axis-based collision detection is applied to allow surface sliding.

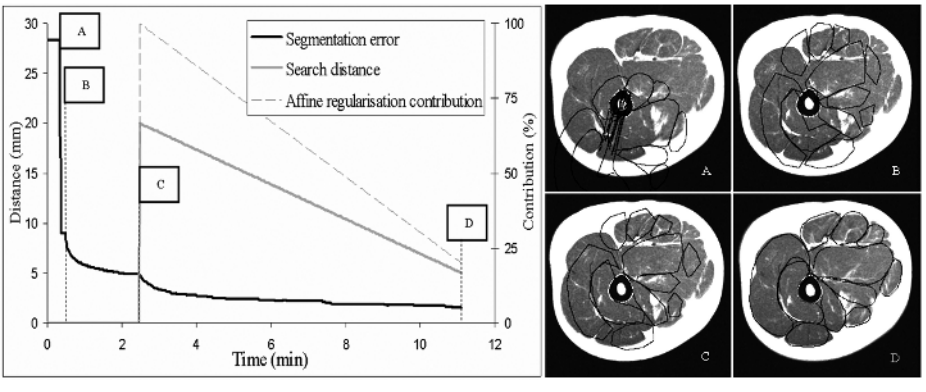


Fig. 2. Automatic muscle segmentation process and result on a sample slice: A) Generic model with wrapped attachments; B) Skeleton-driven model initialization; C) Segmentation using internal forces and skin matching (min. resolution); D) Final segmentation using external forces (max. resolution)

During the segmentation, it is possible to interactively place constraint points on the images, to get a faster matching and a more accurate segmentation. Using our collision detection scheme, deformable models are forced to include or exclude these constraint points. In addition to the generic model, this method has been applied by a medical student to segment accurately 4 datasets from 4 different healthy subjects (2 females and 2 males). We compared automatically segmented models to these reference models (Fig. 2). The average distance (std. dev.) was 1.25mm (1mm) for bones and 1.7mm (1.8mm) for muscles. The overall computation time for the automatic method is around 15min on a standard PC, for which 3/4 of the time is spent on external forces computation. Computation time for each timestep, including visualization, is around 0.5sec.

3.2 Shape Analysis

Musculoskeletal shape characterization is important for anthropometric comparison between individuals, and deformation analysis (temporal and longitudinal studies). For this purpose, our method registers anatomical features through shape and topological constraints. High-level descriptors such as the medial axis convey more information than local descriptors (curvature). Muscle thickness can be simply analyzed through medial axis radii comparison, as shown in Fig. 3b. Using geodesic distances to attachments, we compute normalized coordinates X and Y along medial surfaces, from which a thickness profile can be extracted (maximum radius in Y direction). For some muscles showing thickness steep changes, tendons lengths (which is an important biomechanical parameter) can be automatically extracted (Fig. 3f).

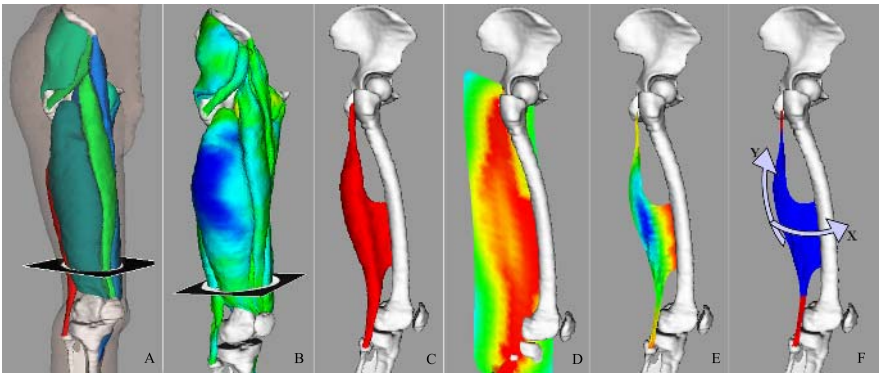


Fig. 3. A) Reconstructed generic model; B) Muscle thickness comparison with an individualized model (blue:≥5mm; red:≤5mm); C) Biceps femoris generic model; D) Initialized medial axis (blue: max radius, red: min radius); E) Medial axis after cropping and fitting; F) Tendon selection (in red) after thickness profile analysis along Y

4 Discussion and Future Work

Prior low-level (e.g. curvature) and high-level (e.g. medial axis) shape information and topological relationships (e.g. proximities, attachments) are relevant for musculoskeletal modelling, and complexity can be efficiently decreased using a multi-resolution approach for force and contact computation. By constraining the problem, our goal is to get a fast and accurate segmentation from a minimum amount of information: we want to extend our previous work on bone motion extraction from real-time dynamic MRI [15] (6 low resolution slices) by extracting soft tissue deformation. We believe that we can get a higher accuracy through a deeper study of intensity profile forces (combination of several metrics, weighting according to profile relevance, etc.) and by adding statistical constraints applied to high-level descriptors (e.g. medial axis radii) according to joint angles. The

next step will be to relate these descriptors with dynamic (e.g. moment arms) and physiological parameters (e.g. muscle activation from EMG). This will provide useful information for validating functional biomechanical models.

Acknowledgments. This work is supported by CO-ME (Computer Aided and Image Guided Medical Interventions) project funded by Swiss National Research Foundation. We would like to thank Dr. Kolo-Christophe, Dr. N'Guyen and Dr. Sadri from the Geneva University Hospital for their collaboration.

References

1. Teran, J., Sifakis, E., Blemker, S., Ng-Thow-Hing, V., Lau, C., Fedkiw, R.: Creating and simulating skeletal muscle from the visible human data set. IEEE TVCG **11** (2005) 317–328
2. Ng-Thow-Hing, V.: Anatomically-based models for physical and geometric reconstruction of humans and other animals. Ph.D. Thesis, Department of Computer Science, University of Toronto (2000)
3. Blemker, S.S., Delp, S.L.: Three-dimensional representation of complex muscle architectures and geometries. Annals of Biomedical Eng. **33** (2005) 661–673
4. Aubel, A., Thalmann, D.: Interactive modeling of the human musculature. Proc. of Computer Animation (2001)
5. Terzopoulos, D., Witkin, A., Kass, M.: Symmetry-seeking models and 3d object reconstruction. International Journal of Computer Vision **1** (1987) 211–221
6. Delingette, H.: General object reconstruction based on simplex meshes. International Journal of Computer Vision **32** (1999) 111–146
7. Montagnat, J., Delingette, H.: 4d deformable models with temporal constraints: application to 4d cardiac image segmentation. MIA **9** (2005) 87–100
8. Baraff, D., Witkin, A.: Large steps in cloth simulation. Proc. of SIGGRAPH98, Computer Graphics **32** (1998) 106–117
9. Volino, P., Magnenat-Thalmann, N.: Implementing fast cloth simulation with collision response. Proc. of the Int. Conference on Computer Graphics (2000) 257–268
10. Teschner, M., Kimmerle, S., Zachmann, G., Heidelberger, B., Raghupathi, L., Fuhrmann, A., Cani, M.P., Faure, F., Magnenat-Thalmann, N., Strasser, W.: Collision detection for deformable objects. Proc. of Eurographics State-of-the-Art Report (2004) 119–135
11. Kaptein, B.L., VanDerHelm, F.C.T.: Estimating muscle attachment contours by transforming geometrical bone models. Journal of Biomechanics **37** (2004) 263–273
12. Pizer, S.M., Fletcher, P.T., Joshi, S., Thall, A., Chen, J.Z., Fridman, Y., Fritsch, D.S., Gash, A.G., Glotzer, J.M., Jiroutek, M.R., Lu, C., Muller, K.E., Tracton, G., Yushkevich, P., Chaney, E.L.: Deformable m-reps for 3d medical image segmentation. International Journal of Computer Vision **55** (2003) 85–106
13. Blum, H.: A transformation for extracting new descriptors of shape. Models for the Perception of Speech and Visual Form (1967)
14. Kalra, P., Magnenat-Thalmann, N., Moccozet, L., Sannier, G., Aubel, A., Thalmann, D.: Real-time animation of realistic virtual humans. Computer Graphics and Applications **18** (1998) 42–56
15. Gilles, B., Perrin, R., Magnenat-Thalmann, N., Vallée, J.P.: Bones motion analysis from dynamic mri: acquisition and tracking. Proc. of MICCAI'04 **2** (2004) 942–949

Towards a Statistical Atlas of Cardiac Fiber Structure

Jean-Marc Peyrat[1], Maxime Sermesant[1], Xavier Pennec[1], Hervé Delingette[1], Chenyang Xu[2], Elliot McVeigh[3], and Nicholas Ayache[1]

[1] INRIA - Asclepios Research Project, Sophia Antipolis, France
jean-marc.peyrat@sophia.inria.fr
[2] Siemens Corporate Research, Princeton, New Jersey, USA
[3] Laboratory of Cardiac Energetics, National Heart Lung and Blood Institute, National Institute of Health, Bethesda, Maryland, USA

Abstract. We propose here a framework to build a statistical atlas of diffusion tensors of canine hearts. The anatomical images of seven hearts are first non-rigidly registered in the same reference frame and their associated diffusion tensors are then transformed with a method that preserves the cardiac laminar sheets. In this referential frame, the mean tensor and its covariance matrix are computed based on the Log-Euclidean framework. With this method, we can produce a smooth mean tensor field that is suited for fiber tracking algorithms or the electromechanical modeling of the heart. In addition, by examining the covariance matrix at each voxel it is possible to assess the variability of the cardiac fiber directions and of the orientations of laminar sheets. The results show a strong coherence of the diffusion tensors and the fiber orientations among a population of seven normal canine hearts.

1 Introduction

While the main geometrical arrangement of myofibers has been known for decades, its variability between subjects and species still remains largely unknown. Understanding this variability is not only important for a better description of physiological principles but also for the planning of patient-specific cardiac therapies [7]. Furthermore, the knowledge of the relation between the myocardium shape and its myofiber architecture is an important and required stage towards the construction of computational models of the heart [14] [1] since the fiber orientation plays a key role when simulating the electrical and mechanical function of the heart. High resolution measurements of fiber orientation has been recently eased with the use of Diffusion Tensor Imaging (DTI) since there is a correlation between the myocardium fiber structure and diffusion tensors [15]. DTI also has the advantage to provide directly this information in 3D with high resolution but unfortunately it is still not available *in vivo* due to the cardiac motion. There has been several studies [12] [13] in the past decade that have measured the variability of fiber orientation from DTI (similar studies has been done for the brain [4]). These studies estimated the fiber direction as

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 297–304, 2006.

the primary eigenvector of each tensor and for instance compared its transmural variation with that observed from dissection experiments.

We propose here to extend these studies by building a statistical model of the whole diffusion tensor and not only its first eigenvector. This tensor analysis allows us to study the variability of laminar sheets which are associated with the tertiary eigenvector. Performing this analysis based on vector analysis (instead of tensor analysis) would have been difficult because the secondary and tertiary eigenvalues have often very similar values and may lead to interpretation errors. To the best of our knowledge, this is the first attempt to perform a first and second order statistical analysis of DT images of canine hearts.

Our statistical analysis proceeds as follows. We first register the canine heart images in a common reference frame using anatomical MRIs. For each heart, we get a deformation field we use to register and transform properly the diffusion tensors considering properties of the cardiac fiber microstructure. Finally we use coherent statistical tools on tensors to study the variability of individual hearts from this average model and to evaluate the relevance of such a model. An application of this framework is carried out using a dataset of seven normal canine hearts.

2 Material and Method

2.1 Data Acquisition

We used a dataset of seven *ex vivo* fixed normal canine hearts acquired [8] and provided by the Center of Cardiovascular Bioinformatics and Modeling (CCBM)[1] at the Johns Hopkins University. Each heart was placed in an acrylic container filled with Fomblin, a perfluoropolyether (Ausimon, Thorofare, NJ). Fomblin has a low dielectric effect and minimal MR signal thereby increasing contrast and eliminating unwanted susceptibility artifacts near the boundaries of the heart. The long axis of the hearts was aligned with the z-axis of the scanner. Images were acquired with a 4-element knee phased array coil on a 1.5 T GE CV/i MRI Scanner (GE, Medical System, Wausheka, WI) using a gradient system with 40 mT/m maximum gradient amplitude and a 150 $T/m/s$ slew rate. Different resolutions have been used around $0.3 \times 0.3 \times 0.9\ mm^3$ and from 14 to 28 gradient directions. The images have been subsampled into $128 \times 128 \times 64$ images with a resolution around $0.6 \times 0.6 \times 1.8\ mm^3$. The temperature during acquisition varied between $18 - 25\,^{\circ}\mathrm{C}$ from one heart to another.

2.2 Myocardium Registration

In order to analyze the statistical variability of the tensors without introducing a bias, we choose to register the images based on independent information: the anatomical MRIs. Before the registration stage, we pre-process semi-automatically the anatomical MRIs by extracting the image background, and by cropping each image above the valve plane. We register each heart on a

[1] http://www.ccbm.jhu.edu/research/DTMRIDS.php

template given by a Procrustes iterative mean estimation. The registration is initialized with an affine global transformation of the hearts defined from three significant landmarks: the apex point and the two corner points of the right ventricle on the valve plane. The apex point is used to define a scaling along the axis of the heart. The matching of the corner points defines a 3D translation, a 2D rotation on the valve plane, and a 2D scaling along a line defined by the 2 corner points. The second registration step is based on a hybrid non-rigid intensity- and landmark-based registration algorithm [2]. This algorithm gives us the ability to interactively refine the registration of a local region. The output of this process is a dense deformation field for each anatomically registered cardiac image.

2.3 Transformation of Cardiac Diffusion Tensors

The next stage is to transform the DT image based on the estimated deformation field. For each voxel, the global deformation field is approximated at the first order by an affine transformation [16]. This underlying affine transformation A is computed from the identity matrix Id and the Jacobian ∇F of the deformation field F: $A = Id + \nabla F$.

Now we have the well known problem of applying an affine transformation to a DT image. Since it was shown that there is a correlation between the cardiac tissue microstructure and the eigensystem of the diffusion tensors [8] [13] [15], we can transform the underlying tissue microstructure with the affine transformation. Then we reconstitute the diffusion tensor from this transformed microstructure knowing the relationship between the two of them. The solution we propose here to handle this problem proved to be similar to the *Preservation of the Principal Direction* (PPD) reorientation strategy proposed in [3]:

$$V_1' = \frac{AV_1}{||AV_1||} \qquad V_3' = \frac{(A^{-1})^T V_3}{||(A^{-1})^T V_3||} \qquad V_2' = V_3 \otimes V_1$$

where V_1, V_2, V_3 are the primary, secondary and tertiary eigenvectors of the original diffusion tensor and where V_1', V_2', V_3' are the ones of the transformed diffusion tensor.

The PPD has been already justified and validated in the case of brain DTI [3] but only partially in the case of cardiac DTI. We propose here a more complete justification of the PPD strategy for cardiac DTI transformation.

As described in Figure 1, the myocardium microstructure is made of laminar sheets of muscle fibers [5]. The space between the laminar sheets is composed of extracellular water and collagen network linking these sheets together. This extracellular space has a mostly unrestricted diffusion in the direction of the two first eigenvectors of the diffusion tensor. The interface plane between the laminar sheets and the extracellular space is an important barrier for water molecules and its normal defines the third eigenvector. The extracellular water between fibers in the laminar planes can explain the difference between the primary eigenvector in the fiber direction and the secondary eigenvector orthogonal to the fiber direction. But the cardiomyocyte geometry is also proposed to explain the privileged diffusion direction in the fiber orientation. A cardiomyocyte is much longer (50-120 μm) than wider (5-25 μm) and much longer than the mean

free path (roughly 10 μm considering the diffusion time of the acquisitions). It means that the cell membrane does not have as much influence on the water diffusion in the fiber direction as in the other directions where the width of the cells is close to the mean free path.

Let us now analyze the effect of the basic transformations (translations, rotations, shears and scaling) describing an affine transformation. The way to transform the fiber structure and thus the diffusion tensors through translation and rotation is obvious. The scaling only changes the density of fibers microstructure inside a voxel considering the acquired voxel resolution and finally not the diffusion rate. The shearing induced by the affine transformation is not so simple to apply to the diffusion tensors. We illustrate in Figure 1 the shearing applied to the basic microstructure of cardiac fibers. The direct transformation of the original eigenvectors V_i leads to the vector AV_i and the transformation deduced from the fiber structure deformation leads to the vector V_i'. As we determine the fiber structure deformation through the fiber direction deformation, the transformed primary eigenvector V_1' is the same as the direct transformation of the original primary eigenvector V_1: $V_1' = \frac{AV_1}{||AV_1||}$.

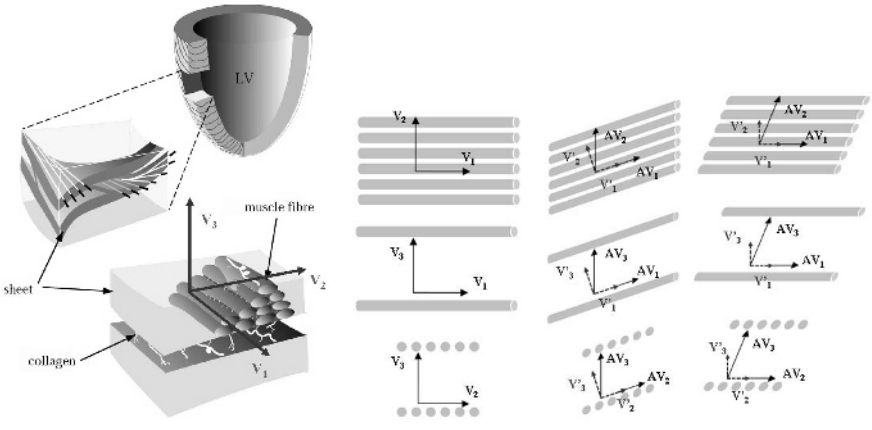


Fig. 1. [Left] Cardiac Fiber Structure (from LeGrice *et al.* [5]). - [Middle Left] Original basic fiber microstructure with eigenvectors V_i. [Middle Right and Right] Shearing applied to the basic fiber microstructure : continuous arrows AV_i are the transformed eigenvectors through the shearing and dashed arrows V_i' are the eigenvectors related to the correlation between the fiber microstructure and the diffusion tensor.

The tertiary eigenvector is defined by n, the vector normal to the laminar sheets which are considered locally plane. The image of a plane through an affine transformation is a plane. It means that these laminar sheets are stable. We just have to determine n' the unit vector normal to the plane defined by the image V_1' of V_1 and the image AV_2 of V_2 [17]: $n' = \frac{(A^{-1})^T n}{||(A^{-1})^T n||}$.

The transformed secondary eigenvector obtained from the correlation between the structure of the fibers happens to be the one that builds an orthonormal basis with the two others. It means that constructing first the secondary (as done for the PPD) or the tertiary (as we do) and then determine the other one to obtain an orthonormal basis leads to the same results. This is the key point that justifies the use of the PPD in our specific case of cardiac DTI.

2.4 Tensor Statistics

The Log-Euclidean framework [10] provides a consistent and rigorous framework to study the statistical variability of DTI for each voxel of the heart. In this framework the space of diffusion tensors is a vectorial space which means it inherits from all the statistical properties and tools we can get from a vectorial space. We thus compute the mean of all the registered DTI and the corresponding covariance [11] at each voxel:

$$\overline{D}_{log} = \exp(\frac{1}{N}\sum_{i=1}^{N}\log(D_i))$$

$$Cov = \frac{1}{N-1}\sum_{i=1}^{N} vect(\Delta D_i)\cdot vect(\Delta D_i)^T$$

where $vect(\Delta D_i)$ is the vectorial representation [11] of $\Delta D_i = \log(D_i) - \log(\overline{D}_{log})$.

The difficulty to visualize the 6×6 covariance matrix of diffusion tensors leads us to study first its norm $\sqrt{Trace(Cov)}$, which allows us to identify the variable and stable regions of the heart. To help us in translating the DTI variability into the fiber structure variability, we can analyze this covariance matrix extracting 6 specific variance parameters at each voxel in the coordinate system of the mean tensor: 3 for each eigenvalue variability and the 3 for each pair of orthonormal eigenvectors orientation variability around the third one [6].

3 Results

We applied the proposed framework to the dataset of seven canine hearts presented previously. We obtain a smooth cardiac DTI atlas catching the shared transmural variation of the fibers directions (see Figure 2) that is similar to the one generally observed [15]. The norm of the covariance matrix in Figure 3 shows a global stability of the compact myocardium and several variable regions especially at the RV and LV endocardial apices where the fiber structure is probably less organized. Some other variabilities at the surface of the heart might also be due to acquisition or registration artifacts.

In order to have a better interpretation of this covariance matrix and to understand the origin of the variabilities, we decompose it into specific modes as described previously. We can see in Figures 3 that the eigenvalues variabilities are homogeneous in the compact myocardium. The mean standard deviations normalized by their mean value $\frac{\overline{\sigma}_i}{\lambda_i}$ are 0.16, 0.24 and 0.22 for the 1^{st}, 2^{nd} and 3^{rd} eigenvalues. Note that these variabilities and their difference are artificially

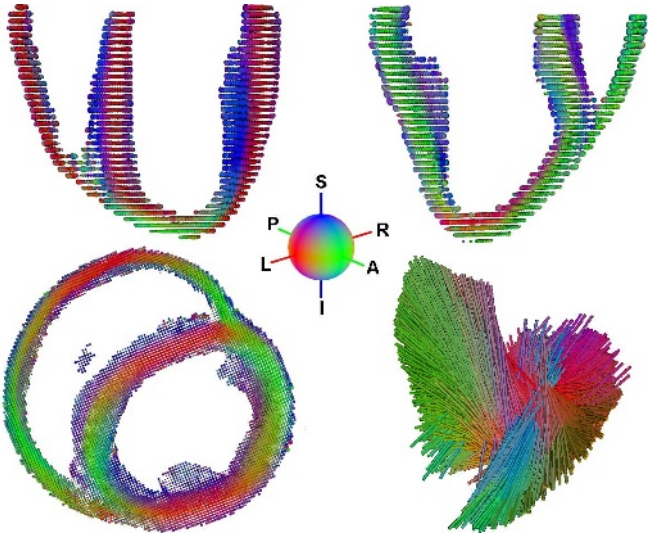


Fig. 2. [*Left and Upper Right*] Average canine cardiac DTI. [*Lower Right*] Fiber tracking in the left ventricle wall with high stiffness parameters for a better visualization of the transmural variation of fibers directions. The RGB colors represent the components of the primary eigenvector $Red = |V_x|$, $Green = |V_y|$, $Blue = |V_z|$ with x, y in the axial plane and z orthogonal to the axial plane as described with the colored sphere.

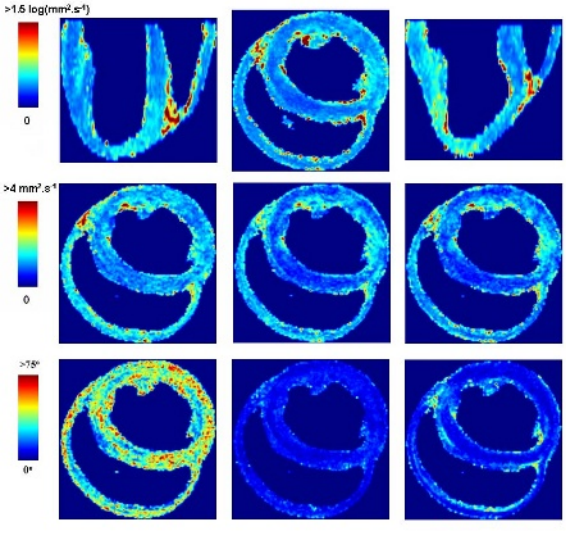


Fig. 3. [*upper*] Norm of the covariance matrix in 3 orthogonal views. [*from middle left to middle right*] Decomposition of the covariance in 6 eigenmodes describing the variability of the 1^{st}, 2^{nd}, 3^{rd} eigenvalues. [*from lower left to lower right*] Rotation variability of the plane orthogonal to the 1^{st}, 2^{nd} and 3^{rd} eigenvectors.

increased by the difference of the temperature during the different image acquisitions (temperature ranging from 18 to 25 °C). Anyway mostly the order of the eigenvalues is really meaningful in terms of fiber architecture and thus in terms of electrical conductivity that are introduced in electromechanical modeling (it does not seem to be any direct correlation between the diffusion and electrical conductivity rates). At least these statistical results give us an idea of the average diffusion rates that could help in detecting locally defined pathologies.

The separation of the eigenvalues variability and the eigenvectors orientation variability is important to evaluate the variability of the myocardial fibers architecture. As seen in Figures 3, the orientation of the fibers is stable among a population (mean standard deviation of 8.8 and 9.4 degrees around the secondary and the tertiary eigenvectors) for the two rotations in the planes containing the primary eigenvector. It means that the fiber orientation of the average cardiac DTI is shared by the dataset. The orientation of the laminar sheets described by the rotation of the plane $Span(V_2, V_3)$ around V_1 shows a much higher mean standard deviation of 23 degrees. First of all this variability is underevaluated considering that the statistical study we proposed is a simplification at the first order [6]. Secondly this variability is due to the difficulty to differentiate the secondary and tertiary eigenvectors when they have similar eigenvalues. When we study the fiber organization (i.e. the 3 rotation eigenmodes) we are not anymore in the diffusion tensor space and an isotropic plane of diffusion leads to a low accuracy in translating it into structural information. Another explanation might also be the existence of two populations of symmetric laminar sheets organization in canine hearts [8] corresponding to the optimal configurations to maximize the systolic shear [9]. Mixing these two populations leads to the computation of an average cardiac fiber architecture model that does not represent a real case of laminar sheets organization. A computation of the normal of the laminar sheets and a further study of their orientation variability is essential to determine its origin and of course to translate an atlas of cardiac DTI into an atlas of the cardiac fiber architecture.

4 Discussion

We proposed a theoretically grounded, simple and powerful framework working directly on diffusion tensors to build an average model and to study their variability over a population of cardiac DTIs. The use of these average models instead of analytical models is an important stage to refine the electromechanical modeling with more accurate and reliable data and also with additional informations concerning the anisotropy in the plane orthogonal to the fiber directions.

This framework is a first step towards a statistical atlas of the cardiac fiber architecture that will probably lead to a better understanding of the cardiac fiber architecture shared by a population of healthy or failing hearts, or to compare and to differentiate populations of hearts (canine-human or normal-failing).

In the next step of our research, a larger dataset of hearts would be helpful to get a statistically more reliable atlas. Also, we would like to apply this framework to human hearts in the context of specific clinical applications.

Acknowledgments

This research was funded by Siemens Corporate Research, Princeton, NJ. Acquisition of the DTI data was funded by the Intramural Research Program of the National Heart Lung and Blood Institute (E.R. McVeigh Z01-HL4004609). We thank Drs. Patrick A. Helm and Raimond L. Winslow at the Center for Cardiovascular Bioinformatics and Modeling for provision of data, A. Azar for his technical support on the registration tools, P. Fillard for provision of diffusion tensors and fiber tracking computation and visualization tools [2], and I.J. LeGrice for provision of an illustration in Figure 1.

References

1. N. Ayache, editor. *Computational Models for the Human Body.* Handbook of Numerical Analysis. Elsevier, 2004.
2. A. Azar et al. An interactive Intensity- and Feature-Based Non-Rigid Registration Framework for 3D Medical Images. In *Proc. of ISBI'06*, 2006.
3. D.C. Alexander et al. Spatial Transformations of Diffusion Tensor Magnetic Resonance Images. *IEEE TMI*, 20(11):1131–1139, 2001.
4. D.K. Jones et al. Spatial normalization and Averaging of Diffusion Tensor MRI Data Sets. *NeuroImage*, 17:592–617, 2002.
5. I.J. LeGrice et al. Laminar structure of the heart: ventricular myocyte arrangement and connective tissue architecture in the dog. *Am J Physiol*, 1995.
6. J.M. Peyrat et al. Towards a statistical atlas of cardiac fiber architecture. Research Report 5906, INRIA, May 2006.
7. M. Sermesant et al. Simulation of cardiac pathologies using an electromechanical biventricular model and XMR interventional imaging. *Med. Image Anal.*, 5(9):467–80, October 2005.
8. P. Helm et al. Ex Vivo 3D Diffusion Tensor Imaging and Quantification of Cardiac Laminar Structure. *Magn. Reson. Med.*, 54(4):850–859, 2005.
9. T. Arts et al. Relating Myocardial Laminar Architecture to Shear Strain and Muscle Fiber Orientation. *Am J Physiol*, 280:H2222–H2229, 2001.
10. V. Arsigny et al. Fast and Simple Calculus on Tensors in the Log-Euclidean Framework. In *Proc. of MICCAI'05*, volume LNCS 3749, pages 115–122, 2005.
11. X. Pennec et al. A Riemannian Framework for Tensor Computing. *IJCV*, 2006.
12. Y. Cao et al. Large Deformation Diffeomorphic Metric Mapping of Fiber Orientations. In *Proc. of ICCV'05*, pages 1379–1386, 2005.
13. P. Helm. *A Novel Technique for Quantifying Variability of Cardiac Anatomy: Application to the Dyssynchronous Failing Heart.* PhD thesis, JHU, 2005.
14. F. Sachse, editor. *Computational Cardiology - Modeling of Anatomy, Electrophysiology, and Mechanics.* LNCS. Springer, 2004.
15. D.F. Scollan. *Reconstructing the Heart: Development and Application of Biophysically-based Electrical Models of Propagation in Ventricular Myocardium Reconstructed from Diffusion Tensor MRI.* PhD thesis, JHU, 2002.
16. R. Sierra. Nonrigid Registration of Diffusion Tensor Images. Master's thesis, ETH Zurich, March 2001.
17. K. Turkowski. *Graphics Gems*, chapter Properties of Surface-Normal Transformations. Academic Press, Inc., 1990.

[2] http://www-sop.inria.fr/asclepios/personnel/Pierre.Fillard/softwares/

A Comparison of Needle Bending Models

Ehsan Dehghan, Orcun Goksel, and Septimiu E. Salcudean

Department of Electrical and Computer Engineering,
University of British Columbia, Vancouver, Canada
{ehsand, orcung, tims}@ece.ubc.ca

Abstract. Modeling the deflection of flexible needles is an essential part of needle insertion simulation and path planning. In this paper, three models are compared in terms of accuracy in simulating the bending of a prostate brachytherapy needle. The first two utilize the finite element method, one using geometric non-linearity and triangular plane elements, the other using non-linear beam elements. The third model uses angular springs to model cantilever deflection. The simulations are compared with the experimental bent needle configurations. The models are assessed in terms of geometric conformity using independently identified and pre-identified model parameters. The results show that the angular spring model, which is also the simplest, simulates the needle more accurately than the others.

1 Introduction

The insertion of needles into soft tissue is an essential part of many medical interventions such as biopsy and brachytherapy. Due to tissue deformation and needle bending during insertion, target displacement, and insufficient feedback from medical imaging, reaching a target location with the needle requires significant skill, training, and experience of the performing physician. Therefore, physical simulators and path planning systems for needle insertion are areas of need for accurate models of flexible needles.

Due to the complexity of interactions between needles and tissue, needle insertion simulation cannot be accommodated using a combined mesh. Therefore, two separate models for the tissue and the needle are usually employed [1,2,3,4] .

Tissue deformation and target displacement during rigid needle insertion have been studied in 2D [1, 2]. Prior work on flexible needle simulation generally assumed needle bending in 2D. DiMaio and Salcudean [2] simulated the needle as an elastic material using the finite element method (FEM) with geometric non-linearity. This method was extended to 3D using 4-node tetrahedral elements by Goksel *et al.* [4] .

Webster *et al.* [3] identified a non-holonomic model for a *highly* flexible needle to simulate the effect of bevel tip on needle motion in tissue. The assumption of high flexibility is unfortunately not valid for many conventional needles such as the ones used for epidural blocks and prostate brachytherapy. This model was later used in [5, 6] for path planning of a needle in 2D tissue. Glozman

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 305–312, 2006.

and Shoham [7] used linear beam elements to simulate needle bending for needle steering. This approach is relatively simple and fast but the elements are not rotation invariant and do not preserve needle length during large deformations. Friction and cutting forces on the needle during insertion have also been studied [8, 9].

Because they are easy to simulate, angular springs have been widely used in cloth and hair modeling [10] and 3D modeling of vibration and deformation [11]. Most of these contributions used complex dynamic models in 3D with various approximations to angular spring behavior (i.e. linear or quadratic). This paper focuses on the physical validity and the modeling accuracy of this method for static deformations of thin rods, and specifically for the deformation of a brachytherapy needle. Different models are assessed in terms of modeling accuracy. The model simulation results are compared with the experimental data. Although the models in this paper are described and validated only in 2D and with a single tip force, they can be easily extended to 3D with multiple forces along the needle. The following section introduces the three different models studied.

2 Methods

2.1 Finite Element Method Using Triangular Elements

The finite element method with triangular elements can be used to predict needle bending. If the needle elastic material is assumed to be linear (mechanical linearity), the linear geometry assumption leads to a set of linear algebraic equations between nodal forces and displacements. While linearity can be exploited for real time implementation, the linear geometry assumption is not valid for needle deformation, due to large rotations and displacements [2].

Using the non-linear geometric relation between strain and displacement leads to a set of non-linear algebraic equations

$$\underline{f} = \hat{\mathbf{K}}(\underline{u})\,\underline{u} \tag{1}$$

where $\underline{u}$ and $\underline{f}$ are the vectors of nodal displacements and forces, respectively. This non-linear formula can compensate for axial displacements and preserves the needle length during large lateral deformations. Although computationally expensive compared to the solution of the linear model, the Newton-Raphson method [12] can be employed to solve the set of non-linear algebraic equations in (1). Simulations show that this method is stable and converges in a few iterations. However, offline computations and condensation cannot be used.

2.2 Finite Element Method Using Non-linear Beam Elements

Another choice of element suitable for needle modeling is the Euler-Bernoulli beam element, which employs linear and cubic interpolation functions for the axial displacement u and the transverse displacement w, respectively. The nodal

variables at node p are the longitudinal displacement u^p, the transverse deflection s_1^p and its derivative s_2^p, such that $s_1^p = w$ and $s_2^p = -dw/dx$. Considering non-linear geometry, the equilibrium equation can be written as [12]:

$$\begin{bmatrix} \mathbf{K}^{11}_{(2\times 2)} & \mathbf{K}^{12}_{(2\times 4)} \\ \mathbf{K}^{21}_{(4\times 2)} & \mathbf{K}^{22}_{(4\times 4)} \end{bmatrix} \begin{bmatrix} \underline{u} \\ \underline{s} \end{bmatrix} = \begin{bmatrix} \underline{f}_1 \\ \underline{f}_2 \end{bmatrix} \tag{2}$$

where $\underline{f}_1$ is the vector of axial forces and $\underline{f}_2$ is the vector of lateral forces and torques on the nodes. $\mathbf{K}^{ij}$ is a function of the material and geometric properties of the beam and the nodal displacements and rotations. This non-linear set of equations can be easily solved using the Picard iterative method [12]. Reduced integration should be used in integrating the stiffness matrix in (2) to avoid *membrane locking* in the element due to large deformations [12].

The linear geometry assumption yields the linear beam element formulation, which separates displacements in axial and transverse directions. This model approximates the transverse deflections with a simple third order curve. Although the linear beam element stiffness matrix is constant and makes offline computation possible, it cannot preserve the needle length during deformation.

2.3 Angular Spring Model

This method models the cantilever using a number of rigid rods connected via rotational springs (see Fig. 1(right)). Consider a short section of length L of a bent cantilever under a constant bending moment M (all in 2D). Fig. 1(left) shows the neutral axis of the bent cantilever and its bending radius ρ. Let c be the distance of the neutral axis from the outer fibre. From similar triangles we obtain the following:

$$\frac{L+\delta}{L} = \frac{\rho + c}{\rho} \Rightarrow \frac{\delta}{L} = \frac{c}{\rho} \tag{3}$$

Substituting (3) in the definition of outer fibre strain and using the fact that $L = \theta\rho$ leads to:

$$\frac{\delta}{L} = \epsilon = \frac{Mc}{EI} \Rightarrow \frac{1}{\rho} = \frac{M}{EI} = \frac{\theta}{L} \tag{4}$$

where E is the stiffness and I is the moment of inertia. With respect to the initial unbent orientation, the angle α of deflection caused by the bent section being

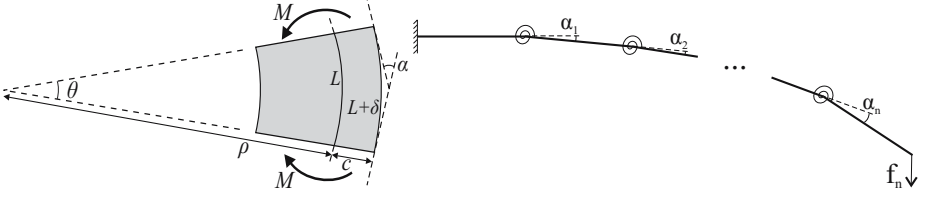


Fig. 1. A short section of a bent cantilever (left), and (right) The angular spring model for a needle

analyzed is equal to θ. Thus, using (4) the linear relation between the bending moment and the deflection angle is written as $M = k\alpha$ where $k = 2EI/L$. While the definition of k obtained for an infinitesimal length may not hold for the discretized model described next, continuum analysis leads to an expected linear relation between M and α.

The angular spring model is composed of line segments connected via springs. In a model with $n+1$ segments the equation between forces and joint angles is:

$$\left[J_1^T(\alpha_1) \,|\, J_2^T(\alpha_1, \alpha_2) \,|\, \cdots \,|\, J_n^T(\alpha_1, \cdots, \alpha_n)\right] \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix} = \mathbf{K} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \tag{5}$$

where α_i are the joint angles, J_i^T is the transposed conventional serial link Jacobian matrix up to link i, f_i is the nodal force on the node $(i+1)$ and $\mathbf{K}$ is a diagonal matrix of spring constants. Since the material is the same throughout the shaft, all springs are assigned the same constant. Despite their non-linearity, these equations can be solved iteratively by fast and simple forward vector-algebraic equations. If the needle is in a posture with values $(\alpha_1, \cdots, \alpha_n)$ at the t^{th} iteration, the next values for the angles are:

$$\alpha_i^{t+1} = \sum_{j \geq i}^{n} M_{ij}^t / k \tag{6}$$

where M_{ij}^t is the torque applied by force f_j on joint i in t^{th} iteration. The procedure converges fast, typically in 2 to 5 iterations.

3 Experiment and Simulations

In order to compare the models with the actual needle bending, an 18 gauge 20 cm brachytherapy needle (Bard, NJ, USA), consisting of a steel stylet sliding inside a steel cannula, was used in an experiment. For simplicity the needle is modeled as a simple bar with a single unknown parameter that is identified by fitting the model to experimental data acquired by bending the needle with a force at the tip. The Young's modulus for the finite element model and the spring constant for the angular spring model are the unknown model parameters to be identified. Since this simplified model represents the combination of the cannula and stylet, the identified Young's moduli are not equal to the Young's modulus of steel and the identified spring constants do not satisfy the definition of k in Section 2.3.

To identify the model parameters, two cost functions of the lateral tip position error and the area error between the models and the real needle were minimized independently. The former is the vertical difference between the simulated and the experimental tip position. The latter is the integral area lying between the simulated and the experimental needle shafts. The identified parameters were used in the model to simulate the bending of the needle.

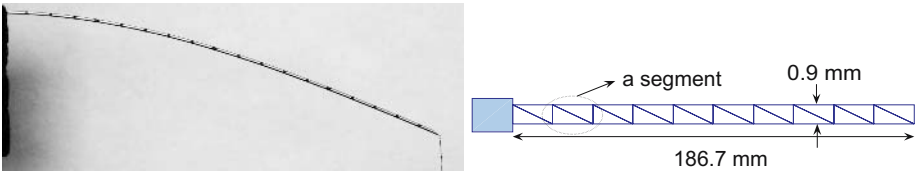


Fig. 2. Left: Needle shaft with the scales hanging at the tip. Right: Discretized needle using a 11×2 nodes mesh.

Table 1. Tip deflection for various lateral tip forces

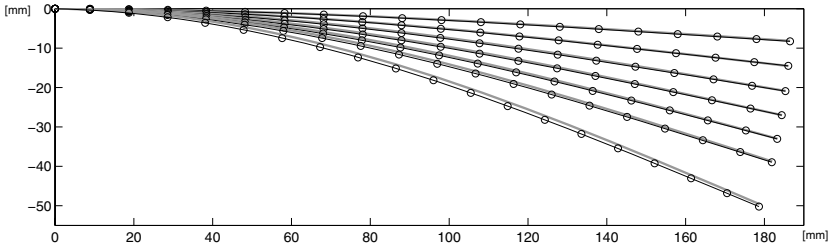
Force [mN]	63	113	163	213	263	313	413
Deflection [mm]	8.3	14.5	20.9	27.0	33.0	39.0	50.2

The brachytherapy needle was horizontally clamped at its base while a vertical force was applied on its tip. The clamp decreased the effective needle length to 18.7 cm. Different vertical tip forces were achieved by hanging different scaling weights of 5, 10, 15, 20, 25, 30, and 40 g on a tiny hook at the tip (see Fig. 2). The stage setup used for hanging weighted 1.3 g. The vertical needle tip force was varied between 63 mN and 413 mN. Images of the needle were taken on a white background with a digital camera, the shutter of which was controlled by computer. The tip deflection from nominal axis is given in Table 1 for each applied force.

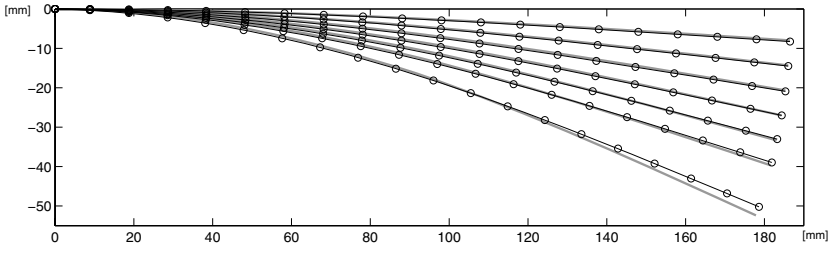
For FEM with triangular elements, meshes of 11×2 and 21×2 nodes were used to discretize the needle as shown in Fig. 2. Plain stress analysis was performed with a thickness of 0.9 mm, and taking the Poisson's ratio as 0.3, equal to the Poisson's ratio of steel. The Young's modulus was identified to fit this model to the experimental data. Similarly, 10 and 20 segment models were used for both the non-linear beam element and the angular spring models. The differences between identified parameters from two different optimization cost functions were less than 2%. Therefore, throughout the rest of this paper we are referring to the values that were identified using the lateral tip-error minimization.

In the triangular FEM and the beam models, the mean values of the identified Young's moduli for different forces were used for simulation and verification. For the triangular element model, this mean value increased with the number of segments. Since the strain is assumed to be constant inside this element a lower Young's modulus (softer material) is needed to yield the same amount of deformation when fewer elements are used. The identified Young's moduli for the non-linear beam element model were constant for any number of elements.

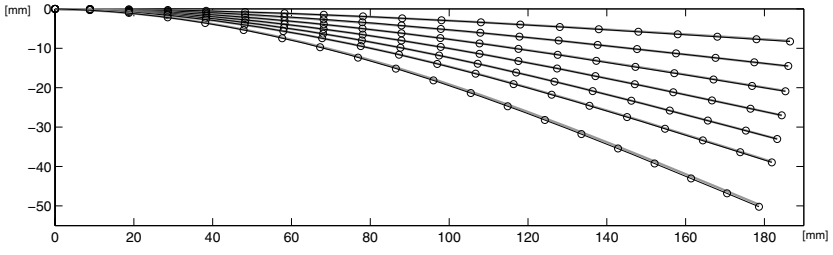
Using the angular spring model, the mean value of the spring constants increased with the change in the number of segments from 10 to 20 and 50 as expected. Normalizing these values with segment length did not show a constant number. In this model the spring constants were identified directly and then used in the simulation phase.



(a) Triangular element with 21×2 nodes.



(b) The non-linear beam element with 20 segments.



(c) Angular spring model with 20 segments.

Fig. 3. The experimental data (circles connected with dashed lines) and the simulated needle (solid lines) using the mean value of identified parameters

In the next step, the needle bending was simulated for a range of tip loads using the mean value of the identified parameters. The identified parameters for all the models did not deviate from their mean values by more than 4%. Figure 3 shows the simulated and the real needle configuration for the three models. For each simulation a tip and an area error are reported in Fig. 4.

4 Discussion

Comparing the two FEM based models, the triangular element model has a slightly higher accuracy than the non-linear beam model. However, the tip position error is less than the needle diameter in most cases. Although the elements of both models have 6 degrees of freedom, modeling a rod using triangular elements requires more elements than in the beam element approach. This leads

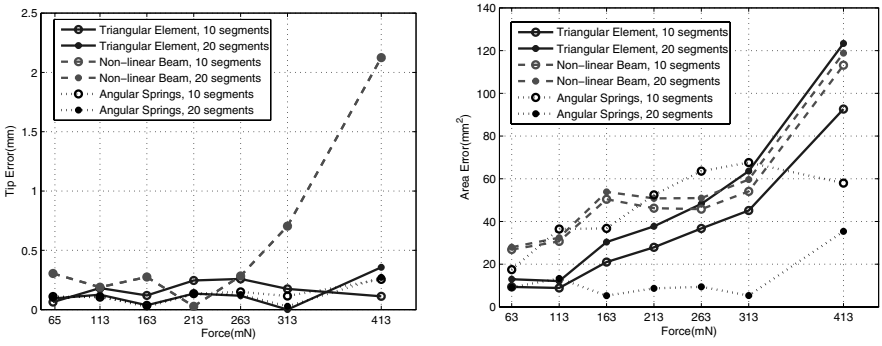


Fig. 4. Comparison of (Left) lateral tip position and (Right) area errors of different models with 10 and 20 segments using the mean value of parameters

to larger matrices. For example modeling a needle using a mesh of $n \times 2$ nodes using triangular elements leads to a $4n \times 4n$ tangent stiffness matrix while using n nodes as beam elements leads to a $3n \times 3n$ one. In addition, the fact that the beam model results are not changing with the number of nodes can be utilized for adaptive meshing in needle insertion simulation to achieve higher speed without loss of accuracy. In adaptive meshing an element is added as the needle penetrates the tissue and makes contact with more tissue mesh nodes [7].

Extension of the non-linear beam model to 3D is straightforward. Assuming that the axial moments and bending are negligible, each element in 3D has 5 degrees of freedom leading to a 10×10 tangent stiffness matrix that can be computed as in the 2D case.

Extension of the triangular element model to 3D leads to tetrahedral elements with 12×12 tangent stiffness matrices. Meshes used in this paper for the triangular element models are asymmetric. For more accurate simulations, symmetric meshes should be used which require more elements and proper design [4]. In contrast to this, the non-linear beam and angular spring models are always symmetric. Therefore, using triangular or tetrahedral symmetric models in 2D or 3D requires a significantly higher number of elements and is computationally more demanding than modeling the needle with non-linear beam elements or with angular spring models.

Of the three models considered, the angular spring model gives the most accurate results, as shown in Fig. 4. It is also computationally more efficient. Extension to 3D is straightforward as presented for hair models in [10].

While not considered explicitly in this paper, the effect of tissue forces that are distributed along the needle, not only at the tip, can be modeled in a similar manner, for both the planar and 3D cases.

5 Conclusion

Three different models for simulating the flexibility of a brachytherapy needle have been demonstrated in 2D. An experimental setup was designed to verify the

models. Model parameters (Young's moduli and spring constants) were identified using the collected data and their mean values were used to simulate needle bending arising from vertical needle tip forces. The fit of the simulations to the experiments was reported in terms of tip error and mismatching area error.

Based on the results one can conclude that the angular spring model is the easiest to implement, the fastest to compute and the most accurate in simulating needle flexibility in 2D.

In the future, the needle bending models will be validated in 3D using the ultrasound images exploiting the methods introduced in [13] .

References

1. Alterovitz, R., Goldberg, K., Pouliot, J., Taschereau, R., Hsu, I.C.: Needle insertion and radioactive seed implantation in human tissue: Simulation and sensitivity analysis. In: Proc. IEEE Int. Conf. Rob. Autom. Volume 2. (2003) 1793–1799
2. DiMaio, S., Salcudean, S.: Interactive simulation of needle insertion models. IEEE Trans. Biomed. Eng. **52** (2005) 1167–1179
3. Webster III, R., Cowan, N., Chirikjian, G., Okamura, A.: Nonholonomic modeling of needle steering. In: Proc. Int. Symp. on Exp. Robotics. (2004) 3337–3343
4. Goksel, O., DiMaio, S.P., Salcudean, S.E., Rohling, R., Morris, J.: 3D needle-tissue interaction simulation for prostate brachytherapy. In: Proc. MICCAI. (2005)
5. Alterovitz, R., Goldberg, K., Chirikijan, G., Okamura, A.: Steering flexible needle under markov motion uncertainity. In: IEEE Int. Conf. Intel. Rob. Sys. (2005) 120–125
6. Alterovitz, R., Goldberg, K., Okamura, A.: Planning for steerable bevel-tip needle insertion through 2D soft tissue with obstacles. In: Proc. IEEE. Int. Conf. Rob. Autom. (2005) 1652–1657
7. Glozman, D., Shoham, M.: Flexible needle steering and optimal trajectory planning for percutaneous therapies. In: Proc. MICCAI. (2004) 137–144
8. DiMaio, S.P., Salcudean, S.E.: Needle insertion modeling and simulation. IEEE Trans. Robotics and Automation **19** (2003) 864 – 875
9. Simone, C., Okamura, A.: Modeling of needle insertion forces for robot-assisted percutaneous therapy. In: Proc. IEEE Int. Conf. Rob. Autom. (2002) 2085–2091
10. Anjyo, K., Usami, Y., Kurihara, T.: Simple method for extracting the natural beauty of hair. Computer Graphics (ACM) **26** (1992) 111–120
11. Anshelevich, E., Owens, S., Lamiraux, F., Kavraki, L.E.: Deformable volumes in path planning applications. In: Proc. IEEE Int. Conf. Rob. Autom. (2000) 2290–2295
12. Reddy, J.N.: An Introduction to Nonlinear Finite Element Analysis. Oxford University Press (2004)
13. Ebrahimi, R., Okazawa, S., Rohling, R., Salcudean, S.: Hand-held steerable needle device. In: Proc. MICCAI. (2003) 223–230

An Inverse Kinematics Model For Post-operative Knee

Ligament Parameters Estimation From Knee Motion

Elvis C. S. Chen[1] and Randy E. Ellis[1,2]

[1] School of Computing, Queen's University, Canada
chene@cs.queensu.ca
[2] Brigham and Women's Hospital, Harvard, U.S.A.
ellis@bwh.harvard.edu

Abstract. A motion-based Inverse Kinematics Knee (IKK) model was developed for Total Knee Replacement (TKR) joints. By tracking a sequence of passive knee motion, the IKK model estimated ligament properties such as insertion locations. The formulation of the IKK model embedded a Forward Kinematics Knee (FKK) [1] model in a numerical optimization algorithm known as the Unscented Kalman Filter [2]. Simulation results performed on a semi-constrained TKR design suggested that ligament insertions could be accurately estimated in the medial-lateral (ML) and the proximal-distal (PD) directions, but less reliably in the anterior-posterior (AP) direction for the tibial component. However, the forward kinematics produced by both the true and estimated ligament properties were nearly identical, suggesting that the IKK model recovered a *kinematically equivalent* set of ligament properties. These results imply that it may not be necessary to use a patient-specific CT or MRI scan to locate ligaments, which considerably widens potential applications of kinematic-based total knee replacement.

1 Introduction

We previously introduced [3] and validated [1] a Forward Kinematics Knee (FKK) model for postoperative knees. Given a set of joint parameters such as ligament insertion locations, our FKK model *predicts* the location of the femorotibial contact for each joint angle using the principle of ligament strain minimization [4]. Knee motion can be reconstructed by finding successive femorotibial contact from full extension to full flexion.

We now introduce an Inverse Kinematics Knee (IKK) model that performs the opposite: by tracking a sequence of knee motion, the IKK model decomposes the motion into a set of *actual* joint angles and the corresponding femorotibial contact locations. The observed joint angles are used by the FKK model to produce a set of *predicted* femorotibial contacts that are contrasted with the observed contacts. Together, the *predictor-actual* pair of contacts are used in the Unscented Kalman Filter [5] to estimate the joint parameters that would lead to the observed knee motion.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 313–320, 2006.

2 Method

Articular surfaces of a size-3 Sigma Knee (Johnson & Johnson) were laser-scanned at a resolution of $0.4mm$, resulting in two point clouds of approximately $31,000$ and $19,000$ points for the femoral (F) and tibial (T) components, respectively. Joint coordinate systems [6] were assigned to the components. The *absolute*, space-fixed coordinate system was associated with the tibial component, and the *relative*, body-fixed coordinate system was associated with the femoral component. Without loss of generality, the Z-axes were aligned with the anatomical axis of the lower limb. The X-axes were perpendicular to the Z-axes lying on the sagittal plane with the anterior direction being positive. The Y-axes were derived as the cross product of the two: $Y = Z \times X$. The centroid of the point clouds were chosen as the origin of the coordinate systems.

These two coordinate systems were related by a homogeneous transformation. If $\bar{p}^f$ is a 3×1 column vector that measures the coordinate of a point in the femoral system, then its corresponding tibial location $\bar{p}^t$ can be expressed as:

$$\begin{bmatrix} \bar{p}^t \\ 1 \end{bmatrix} = \begin{bmatrix} R\ \bar{d} \\ 0\ 1 \end{bmatrix} \begin{bmatrix} \bar{p}^f \\ 1 \end{bmatrix} \tag{1}$$

where R is a 3×3 rotation matrix and $\bar{d}$ is a 3×1 translation vector. Because the tibial was assumed to be fixed, R and $\bar{d}$ represented the relative joint angle and position of the femoral component. The rigid-body transformation of a point is a rotation (R) to the space-fixed coordinate system followed by a linear displacement ($\bar{d}$).

2.1 Forward Kinematics Knee Model

The FKK for the passive knee was developed based on the principle of ligament strain minimization[4]: In a passive knee where no external force and torque is present, the knee would rest into an equilibrium position where the strain stored in the knee ligaments is minimized. The passive FKK model can be formalized as:

$$\bar{d} = FKK(F, T, R, \bar{L}) \tag{2}$$

where F and T are the geometry of the femoral and the tibial TKR components, respectively, R is the femorotibial joint angle, and $\bar{L}$ is a state vector representing the mechanical properties of knee ligaments. $\bar{L}$ is a concatenation of n filament state vectors, each with the following format:

$$\bar{l}^i = [p_x^f, p_y^f, p_z^f, p_x^t, p_y^t, p_z^t, K, B]^T \tag{3}$$

where $\bar{l}^i$ represents the i^{th} filament of the ligament, $[p_x^f, p_y^f, p_z^f]^T$ and $[p_x^t, p_y^t, p_z^t]^T$ are the 3D femoro- and the tibial- insertion locations, respectively, and K and B are the optional spring constants for the filament. The filaments are modeled as springs that store no compressive force, with the strain energy of each filament calculated as:

$$E = \begin{cases} .5 \times K \times (L' - \tilde{L})^2 + B \times (L' - \tilde{L}), & \text{if } L' \geq \tilde{L} \\ 0, & \text{if } L' < \tilde{L} \end{cases} \tag{4}$$

where $\tilde{L}$ is the neutral length of the filament, L' is the Euclidean distance between filament attachment points, and K and B are the spring constants.

The vector $\bar{d}$ in Eqn. (2) represents the femorotibial contact location where the total ligament strain energy stored in $\bar{L}$ is minimized [3]; thus, the 4×4 homogeneous matrix $\begin{bmatrix} R & \bar{d} \\ 0 & 1 \end{bmatrix}$ completely specifies the femorotibial pose. Knee kinematics can be reconstructed by varying R from full extension to full flexion and computing the corresponding displacement vectors $\{\bar{d}\}$.

2.2 FKK Model Validation

A spring-ligament apparatus mimicking passive TKR knee was constructed as a validation tool for the FKK model [1]. The apparatus is composed of the laser-

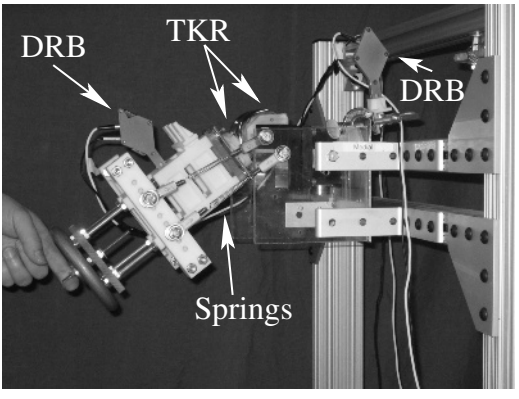


Fig. 1. A spring-ligament apparatus simulating the kinematics of passive knee

scanned TKR components held in contact by the tensile forces stored within a set of mechanical springs. The mechanical properties needed to satisfy Eqn. (3) were determined for each spring.

A Dynamic Reference Body (DRB) was rigidly attached to each TKR component, allowing the motion of the components to be tracked with an optical tracker (OPTOTRAK 3020, NDI, Canada). Let P_f and P_t be the poses of the DRBs, in the OPTOTRAK coordinate system, attached to the femoral and tibial components, respectively, and let Q_f and Q_t be the 4×4 registration matrices that register the scanned femoral and the tibial point clouds to their respected components. Then for a given tracked instance, the relative femorotibial pose, in the scanned tibial point cloud coordinate system, is given by Eqn. (5):

$$\begin{bmatrix} R' & \bar{d}' \\ 0 & 1 \end{bmatrix} = Q_t^{-1} P_t^{-1} P_f Q_f \tag{5}$$

where R' is the *actual* femorotibial angle and the $\bar{d}'$ is the *actual* femoral displacement. Thus, knowing the spring ligament properties of the apparatus, the

observed *actual* femorotibial angle R' can be used as the input to Eqn. (2); the computed virtual displacement vector $\bar{d}$ be compared to the actual displacement vector $\bar{d}'$.

Given accurate inputs needed by Eqn.(2), the FKK model was found [1] to predict the femorotibial contact locations (and the displacement vector $\bar{d}$) with sub-millimeter accuracy. Simulation results suggested that, when a semi-constrained TKR (such as the Sigma) is used, a simple single-bundle ligament model is sufficient to generate accurate knee kinematics: it is better to get a good rough estimate of knee-ligament geometry rather than to toil for a other aspect of a complex ligament model.

2.3 Unscented Kalman Filter

The Square-Root Unscented Kalman Filter [5] (SQ-UKF) was implemented as the numerical optimization algorithm for its numerical stability and efficiency. Kalman and related filters operate in a *predictor-corrector* and *iterative* fashion. In the Kalman filter paradigm, the state of a stochastic process is represented by a set of variables that evolves through time according to a *process* model. The state variables are related to a set of observable variables through a *measurement* model. When an new observation is available, the current state vector is advanced in time through the process model and a *predicted* measurement is generated through the measurement model accordingly. The difference between the actual and the predicted measurements, called the *innovation*, is used to correct the state vector so the state vector better fits the observation.

We apply the UKF to ligament parameter estimation in which the ligament properties were treated as the unknown. The *process* and *measurement* models needed for UKF are:

$$\bar{L}_{k+1} = \bar{L}_k + v_k \tag{6}$$

$$\bar{d}_k = FKK(F, T, R, \bar{L}_k,) + n_k \tag{7}$$

That is, the process model for parameter estimation is the *identity* function, because the quantities to be estimated are assumed to remain constant over time. The measurement model is the FKK model itself, where the quantity to be measured, as a function of the ligament state vector, is the displacement vector $\bar{d}$ of a given femorotibial angle R. The noise vectors v_k and n_k refer to the uncertainties of the process and measurement models, respectively.

A set of 1089 *synthetic* displacement vectors $\bar{d}$ of varying femorotibial angles was precomputed using a 3-ligament (PCL, MCL, and LCL) FKK model. The single-bundle ligament geometry was adapted from our previous study [3]: the single-bundle ligament was artifically generated by taking the geometrical mean of the ligament insertions in the multi-bundle ligament configuration, using the summed spring constants. The femorotibial angle ranged from 0° to 120° in flexion, -1° to 1° in varus, and -1° to 1° in internal angulation, all at 1° intervals. This set of *synthetic* displacement vectors is destinated as the *true* observations since it was calculated using the known and correct ligament geometry.

Due to the computational complexity, our simulations were performed using a 2-ligament state vector (MCL plus LCL) in which the ligament properties of the PCL were assumed known. Furthermore, the spring constants needed for Eqn.(3) were assumed to be known because the formulation of Eqn. (4) did not require the absolute value of the spring constants, only the relative stiffness between the ligaments [7]. Hence the state vector for the IKK model has a dimensionality of 12, in which the femoral and the tibial insertions of both the MCL/LCL were the unknowns.

The iterative process of UKF parameter estimation proceeded as follows. First, an initial guess of each of the ligament insertion was given and the state vector $\bar{L}$ need for Eqn. (2) was constructed according to Eqn. (3). At each itera-

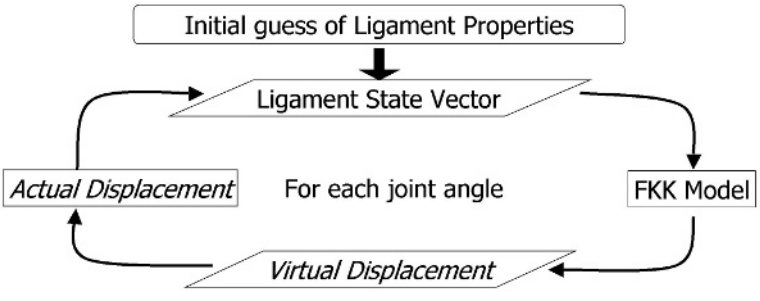


Fig. 2. The iterative process of UKF parameter estimation. When a new observation becomes available, the current estimate of the state vector is used to predict the femorotibial contact locations using the FKK model. The difference in the predicted and the observed femorotibial contact locations, call the *innovation*, is used to correct the state vector estimation.

tion when a new observation (i.e. an actual displacement $\bar{d}'$) became available, a *predicted* measurement (i.e. $\bar{d}$ using Eqn. (2)) was computed using the estimated state vector. The *innovation* term was calculated and propagated through the UKF formulation [5], and the state vector was updated accordingly.

3 Results

Figure 3 depicts a typical result of the ligament parameter estimation using the UKF paradigm. Each of the 4 ligament insertion locations was erroneously and intentionally chosen at a location $10mm$ away from the true insertion.

Figure 3(a) shows the convergence of each of the MCL-femoral, MCL-tibial, LCL-femoral, and LCL-tibial insertion from their respective initial guess. During the UKF iterations, all 4 insertion estimates converged to a steady-state after about 200 observations. After this point, the LCL-tibial insertion estimate remained about $4.0mm$ away from the true location, and the MCL-tibial insertion estimate remained about $4.7mm$ away from the true location. The femoral insertion estimates for both ligaments had errors about $2.0mm$

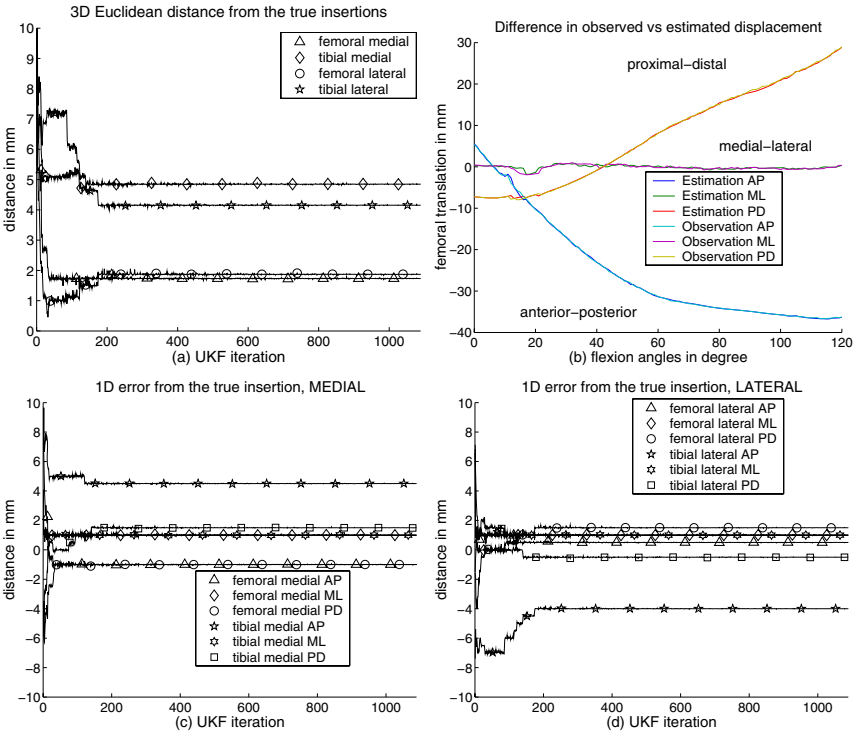


Fig. 3. Simulation results for ligament parameter estimation using UKF. Insertion locations of both MCL and LCL were initialized to be $10mm$ away from the true location.

After the ligament estimates reach a steady state, they were used in the FKK model to produce a sequence of passive knee motion from full extension (0°)to full flexion (120°) at 1° intervals. Figure 3(b) depicts a visualization of knee kinematics by plotting the (x, y, z) components of the displacement vector $\{\bar{d}\}$: Both the true and the estimated ligament insertions produced near identical passive forward kinematics.

Figure 3(c) and (d) show the convergence of the ligament insertions of the MCL and the LCL ligament, respectively, in each of the AP, ML, and PD directions. For MCL ligament estimation (Figure 3(c)), the largest error was the tibial AP (x-axis) direction, which had a magnitude of about $4mm$. Errors in the other two directions had a magnitude of $2mm$ or smaller. For LCL ligament estimation (Figure 3(d)), the largest error was the tibial AP (x-axis) direction, with a magnitude of about $4mm$. Errors in other directions were $2mm$ or smaller in magnitude.

4 Discussion

The complexity of the SQ-UKF parameter estimation algorithm is $O(n^2)$ [5], where n is the dimension of the state vector $\bar{L}$. Using a 2GHz computer with

sufficient memory, the IKK model took about $5-7$ days to iterate through 1089 observations. It should be noted that the IKK process can be terminated as soon as the state vector reaches a steady-state: on the example shown above, the IKK model provided a steady-state estimate in about 1 day. In addition, the UKF can be implemented with ease in a multi-threaded environment: up to $(2n+1)$ computers can be used to obtain a linear speed-up.

In the example shown above, the parameter estimation process was not able to fully recover all components in the state vector. In particular, the tibial anterior-posterior component for both MCL and LCL insertion sites had much higher error compared to other two directions. However, the resulting forward kinematics generated by both the true and the estimated ligament insertions were nearly identical. This explains why the IKK could not converge any further: the *innovation* vectors needed to correct the state vector were almost zero and thus had negligible effect on updating the state vector. One can also think of this phenomena as reaching a local minima, where there exist multiple solutions to the IKK model. Thus, the IKK model has recovered a *kinematics-equivalent* ligament properties.

The sensitivity of the surgical placement on the resulting kinematics can be attributed to the geometry of the TKR components. Most TKR designs are *more* congruent on the coronal plane than on the sagittal plane; thus a small misplacement of the TKR components in the coronal plane would greatly influence the elongation of the knee ligaments (see Eqn. (4)) changing the femorotibial contact location. For the same reason, the surgical placement of the femoral component would have a more profound effect on the knee ligament because is is more convex than the tibial component.

5 Conclusion

We introduced an Inverse Kinematics Knee model that embeds a Forward Kinematics Knee model in the Unscented Kalman Filter paradigm. Given an observed knee motion and an initial guess of the ligament properties, the IKK model recovers a set of *kinematics-equivalent* ligament properties with a semi-constrained TKR design. Simulation results suggested that:

- Surgeons have some freedom in the surgical placement of the tibial component in the anterior-posterior direction without changing the postoperative passive kinematics,
- Precise placement of the femoral component is indicated as the resulting kinematics are sensitive to the placement,
- There is a direct, and yet complex, relationship between the TKR design, the surgical placement, and the resulting knee kinematics: The less congruent the TKR geometry is in a given direction, the greater freedom the surgeon has in the surgical placement of the component in that direction.

This study is limited by the number of ligaments examined, the configuration of the ligaments, and the number of prosthesis designs included in the study. Future

works include applying the IKK model to more TKR designs, for postoperative TKR assessment, and extending the IKK model to natural knees.

We believe that the IKK model may provide a new paradigm for TKR surgical planning; the ligament insertions may be estimated prior to TKR surgery. These results may suggest that it is not necessary to use a patient-specific CT or MRI scan to locate ligaments, which considerably widens potential applications of kinematic-based total knee replacement. Further inquiry into this subject is indicated.

References

1. Chen, E.C.S., Lanovaz, J.L., Ellis, R.E.: Ligament strains predict knee motion after total joint replacement. In Duncan, J., Gerig, G., eds.: Medical Image Computing and Computer-Assisted Intervention - MICCAI 2005. LNCS 3749, Springer-Verlag (2005) 770–777
2. Wan, E.A., van der Merwe, R., Nelson, A.T.: Dual estimation and the unscented transformation. In Solla, Todd K. Leen, and Klaus-Robert Mller, S.A., ed.: Advances in Neural Information Processing Systems 12, MIT Press (2000) 666–672
3. Chen, E., Ellis, R.E., Bryant, J.T., Rudan, J.F.: A computational model of postoperative knee kinematics. Medical Image Analysis **5**(4) (2001) 317–330
4. Essinger, J.R., Leyvraz, P.F., Heegard, J.H., Robertson, D.D.: A mathematical model for the evaluation of the behaviour during flexion of condylar-type knee prostheses. Journal of Biomechanics **22**(11–12) (1989) 1229–1241
5. van der Merwe, R., Wan, E.A.: The square-root unscented kalman filter for the state and parameter-estimation. International Conference on Acoustics, Speech, and Signal Processing 2001, Salt Lake City, Utah (2001)
6. Grood, E.S., Suntay, W.J.: A joint coordinate system for the clinical description of three-dimensional motions: application to the knee. Journal of Biomechanical Engineering **105** (1983) 136–144
7. Martelli, S., Ellis, R.E., Marcacci, M., Zaffagnini, S.: Total knee arthroplasty kinematics: Computer simulation and intraoperative evaluation. The Journal of Arthroplasty **13**(2) (1998) 145–155

Online Parameter Estimation for Surgical Needle Steering Model

Kai Guo Yan[1], Tarun Podder[3], Di Xiao[1], Yan Yu[3], Tien-I Liu[4], Keck Voon Ling[2], and Wan Sing Ng[1]

Schools of [1] MAE, [2] EEE, Nanyang Technological University, Singapore
[3] Department of Radiation Oncology, University of Rochester, NY, U.S.A.
[4] Computer Integrated Manufacturing Lab, California State University, Sacramento, California, U.S.A.

Abstract. Estimation of the system parameters, given noisy input/output data, is a major field in control and signal processing. Many different estimation methods have been proposed in recent years. Among various methods, Extended Kalman Filtering (EKF) is very useful for estimating the parameters of a nonlinear and time-varying system. Moreover, it can remove the effects of noises to achieve significantly improved results. Our task here is to estimate the coefficients in a spring-beam-damper needle steering model. This kind of spring-damper model has been adopted by many researchers in studying the tissue deformation. One difficulty in using such model is to estimate the spring and damper coefficients. Here, we proposed an online parameter estimator using EKF to solve this problem. The detailed design is presented in this paper. Computer simulations and physical experiments have revealed that the simulator can estimate the parameters accurately with fast convergent speed and improve the model efficacy.

Keywords: Online parameter estimation, spring/damper model, EKF.

1 Background

Precise needle insertion is very important for a number of percutaneous interventions. Yet it's very difficult to achieve in real practice e due to tissue heterogeneity and elastic stiffness, unfavorable anatomic structures, needle bending, inadequate sensing, tissue/organ deformation and movement, and poor maneuverability. Several research groups (e.g. [1,2]) have implemented computer and robotic assistance to align the needle with the target assuming known target location in 3D, target immobilization, and a straight needle path. Errors caused by needle deflection and tissue deformation have been observed for a long time. Yet to date, there are few effective physical-based needle steering systems exist for correcting the needle deflection when it's observed. In addition, many procedures are currently not amenable to needles because of obstacles, such as bone or sensitive tissues, which lie between feasible entry points and potential targets. Thus, there is a clear motivation for needle steering in order to provide accurate and dexterous targeting.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 321 – 329, 2006.

To solve this problem, a spring-beam-damper needle steering model has been developed [3], in which the tissue is represented by a series of spring-damper system. The proposed spring-beam-damper model takes into consideration both the linear and nonlinear tissue reaction effects when interacting with the flexible needle. In the literature, the spring-damper model has been adopted by many research groups [4-7] in studying the tissue deformation. One difficulty in using such model is to estimate the spring and damper coefficients of the tissue, especially when applying to a human in a real medical surgery.

This paper aims to design an effective parameter estimator for the spring-damper model we have developed. Considering the steering model is nonlinear, time varying, and at the same time, the dataset collected is very noisy, here we proposed an online parameter estimator using EKF to estimate the coefficients in real time. The detailed design is presented in this paper. Extensive computer simulations have been carried out to assess the efficacy of the proposed estimator. Results and conclusions are presented in the following sections.

2 Design of the Online Parameter Estimator

2.1 The Dynamic Equations of Needle Steering Model

This section briefly reviews the spring-beam-damper model we have proposed (refer to [3] for more details). Fig. 1 shows the mechanism of the needle steering procedure. The flexible needle is assumed to follow the Bernoulli-Euler beam model and to be clamped tightly at end. The tissue is represented by a series of spring-damper system. The initial lengths of springs are decided by the needle tip trajectory. At the beginning, the needle is placed next to the tissue. As the needle inserts into the tissue, the springs and dampers come into contact with the needles body and exert force on it. The forces of the springs at time instant t are determined by the needle body shape at that time and the forces of the dampers are determined by the velocities of the contact points.

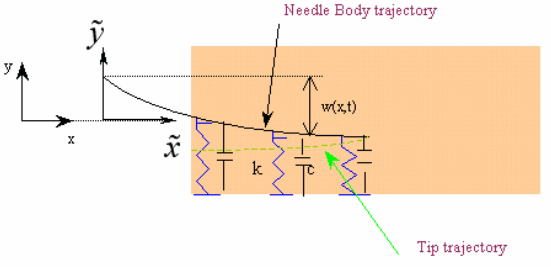


Fig. 1. Mechanism of the needle steering procedure

The system dynamic equation can be derived using Hamilton's principle as follows:

$$\int_{t_1}^{t_2} (\delta T - \delta V + \delta W_{nc}) dt = 0 \tag{1}$$

with *T, V, W_{nc}* representing the kinetic energy, potential energy and work done by non-conservative forces respectively.

Unconstrained modal analysis is adopted in solving the partial differential equations. By unconstrained modal analysis [8], the deflection of the beam at $\tilde{x}$, ω($\tilde{x}$,t) and the position of the fixture, y(t) can be represented, respectively, as

$$y = \alpha(t) + \beta q(t) \tag{2}$$

$$\omega(\tilde{x}, t) = \phi(\tilde{x}) q(t) \tag{3}$$

where α(t) describes the motion of the center of mass of the total system without perturbation, $\phi(\tilde{x})$ is the shape function, $q(t)$ is the time-varying amplitude of motion. β is defined to satisfy

$$\beta = -\frac{\rho}{M_t} \int_0^L \phi(\tilde{x}) d\tilde{x} \tag{4}$$

This allows transforming the nonhomogeneous equations into a set of second-order ordinary differential equations. After some algebraic manipulations, the model is finally derived as,

$$\begin{aligned} &\dot{\alpha}(t) \int_{L-v_x t}^{L} c\varphi_i(\tilde{x}) d\tilde{x} + \alpha(t) \int_{L-v_x t}^{L} k_0 \varphi_i(\tilde{x}) d\tilde{x} + \ddot{q}_i(t) + \sum_j \dot{q}_j(t) \int_{L-v_x t}^{L} c\varphi_i(\tilde{x}) \varphi_j(\tilde{x}) d\tilde{x} \\ &+ \sum_j q_j(t) \int_{L-v_x t}^{L} k_0 \varphi_i(\tilde{x}) \varphi_j(\tilde{x}) d\tilde{x} + \omega_i^2 q_i(t) = F_y \beta_i + f_{1i} \end{aligned} \tag{5}$$

$$\begin{aligned} &M_t \ddot{\alpha}(t) + c v_x \dot{\alpha}(t) t + \sum_j \dot{q}_j(t) \int_{L-v_x t}^{L} c\varphi_j(\tilde{x}) d\tilde{x} + \sum_j q_j(t) \int_{L-v_x t}^{L} k_0 \varphi_j(\tilde{x}) d\tilde{x} + k_0 v_x \alpha(t) t \\ &= F_y + f_2 \end{aligned} \tag{6}$$

where,

$$M_t = M + \rho L \tag{7}$$

$$\begin{aligned} f_{1i} &= \int_0^{t_1} k_0 \alpha(t_1) \varphi_i(v_x t_1 + L - v_x t) v_x dt_1 + \int_0^{t_1} k_0 \sum_j \varphi_j(L) q_j(t_1) \varphi_i(v_x t_1 + L - v_x t) v_x dt_1 \\ &= \int_0^{t_1} k_0 Y(t_1) \varphi_i(v_x t_1 + L - v_x t) v_x dt_1 \end{aligned} \tag{8}$$

$$f_2 = \int_0^{t_1} k_0 \alpha(t_1) v_x dt_1 + \int_0^{t_1} k_0 \sum_j \varphi_j(L) q_j(t_1) v_x dt_1 = \int_0^{t_1} k_0 Y(t_1) v_x dt_1 \tag{9}$$

here, M is the mass of the fixture, which links the needle with the 3D motion platform. M_t is the total mass of the fixture and needle. Y(t1) is the needle tip position at time instant t1. F_y is the steering force, which is acted at the needle end in y direction. V_x is needle insertion speed. L is the length of the elastic beam. *i* and *j* here refer to different mode shape. We consider continuous distributed spring and dampers along the needle body. So here k_0 is stiffness coefficient of the spring per unit length, and c is the damping coefficient per unit length. At current stage, k_0 and c are assumed to be constant.

2.2 Design of the Parameter Estimator

Estimation of the system parameters, given noisy input/output data, is a major field in control and signal processing. Many different estimation methods have been proposed in recent years [9-13]. Among various methods, Extended Kalman filtering (EKF) is very useful for estimating the parameters of a nonlinear and time-varying system. And at the same time, it can remove the effects of interferences and noises to achieve satisfactory results. Considering the needle steering model is nonlinear and time varying, and furthermore, the data we obtain are usually very noisy, we adopt EKF method to design the parameter estimator.

The steering model is rearranged in the state-space form. Here the model spring and damper coefficients are treated as states too. For simplicity, we express the needle steering model as

$$\dot{x} = g(x) + w \tag{10}$$

w is a random zero-mean process. The measurement equation, the output tip position is a linear function of the states. Because the measurements are discrete, we write the equation as

$$y_k = cx_k + v \tag{11}$$

where $c = [1 \quad \phi(L) \quad 0 \quad 0 \quad 0 \quad 0]$ and v is a zero-mean random process.

Because the system equations are nonlinear, a first-order approximation is used in the continuous Riccati equations for the systems dynamics matrix F. The matrices are related to the nonlinear system according to

$$F = \left.\frac{\partial g(x)}{\partial x}\right|_{x=\hat{x}} \tag{12}$$

The fundamental matrix, required for the discrete Riccati equations, can be approximated by the Taylor-series expansion. In this application, the fundamental matrix will only be used in the calculation of the Kalman gains. As demonstrated in [14] that adding more terms to the Taylor-series expansion for the fundamental matrix may not improve the performance of the overall filter, here only the first two terms are used to approximate the series.

$$\Phi_k \approx I + FT_s \tag{13}$$

The Riccati equations required for the computation of the Kalman gains, are [14]

$$\begin{aligned} M_k &= \Phi_k P_{k-1} \Phi_k^T + Q_k \\ K_k &= M_k C^T (C M_k C^T + R)^{-1} \\ P_k &= (I - K_k C) M_k \end{aligned} \tag{14}$$

here, Φ_k is the fundamental matrix, P_k is a covariance matrix representing errors in the state estimates after an update and M_k is the covariance matrix representing errors in the state estimates before an update. C is the coefficient matrix of output. Because Φ_k is nonlinear function of the state estimates, the Kalman gains cannot be computed offline as is possible with the linear kalman filter. Q_k and R are the process-noise matrix and measurement noise matrix, respectively. The discrete process-noise matrix Q_k can be found from the continuous process-noise matrix according to

$$Q_k = \int_0^{Ts} \Phi(\tau) Q \Phi^T(\tau) dt \tag{15}$$

With an extended Kalman filter, the new state estimate is the old state estimate projected forward to the new sampling or measurement time plus a gain times a residual or

$$\tilde{x}_k = \bar{x}_k + K_k [y_k - \bar{y}_k] \tag{16}$$

here K_k is the Kalman gain, y_k is the measured output and $\bar{y}_k$ is the estimated output. The old estimates that have to be propagated forward do not have to be done with the fundamental matrix but instead can be propagated directly by integrating the actual nonlinear differential equations forward at each sampling interval.

3 Validation of the Parameter Estimator

This parameter estimator has been implemented using Matlab 7.0. To validate whether this estimator can estimate the parameter fast and accurately, firstly we used the pre-programmed model generator to generate the output needle tip position data with input force 10^{-4} (N), the spring and damper coefficients per length 2 (N/m^2) and 5 (N.s/m^2), respectively. Then these input and output data were filled into the parameter estimator. The initial states are chosen to be [10^{-7}; 10^{-7}; 10^{-7}; 10^{-7}; 1; 4]. The initial covariance matrix is estimated to be diag([0 0 0 0 4e+5 5e+5]). The simulation results have been presented in Fig. 2 and Fig. 3. The RMS error for output is 6.9888e-006. We can see that the estimator can estimate the coefficients accurately with fast convergent speed.

The initial covariance matrix is very important for the convergent rate. Reasonable chosen initial covariance can improve the convergent rate, as shown in Fig. 4. Generally speaking, the convergent rate is increased with larger covariance matrix. But if the covariance matrix is set to be too large, the estimation will fluctuate and overshoot, as shown in Fig. 4. The setting of the covariance matrix is rather rule-of-thumb. It has been found during experiments that, for the last two terms in the initial covariance matrix, the magnitudes of 10^5 can always give good convergence rate. Exceeding this magnitude may result in fluctuation and overshoot.

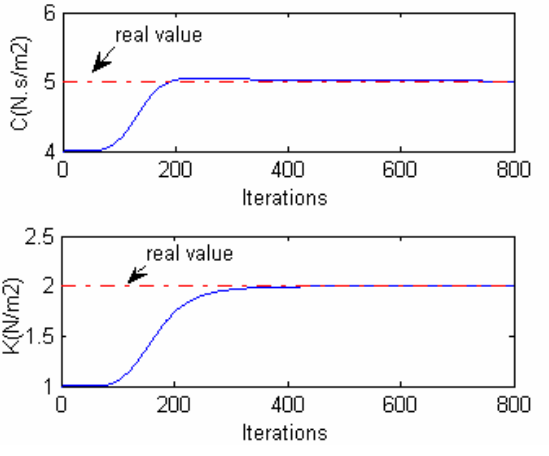


Fig. 2. Estimated damper & spring coefficient vs. real value

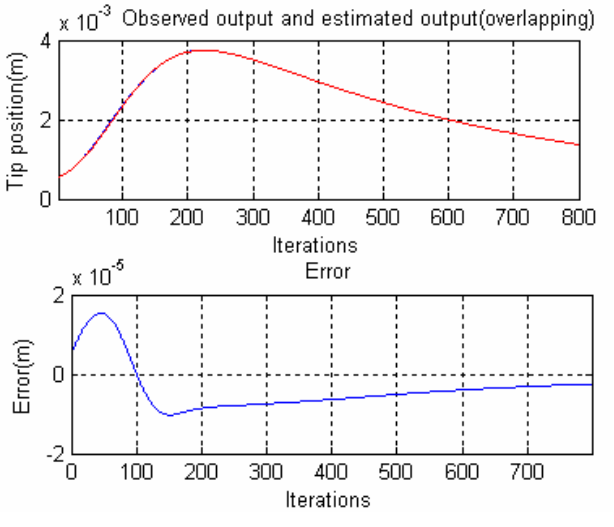


Fig. 3. Estimated needle tip position vs. measured tip position

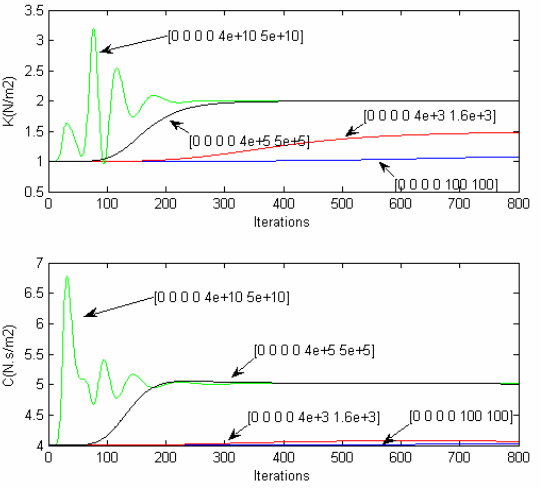


Fig. 4. Coefficient estimation comparison with different covariance matrixes

4 Validation of the Steering Model Using Parameter Estimator

The design purpose of this parameter estimator is to validate the parameters of the needle steering model. These parameters will be set into the steering model to do the simulation. The simulated output will then be compared with the measured output to validate the effectiveness of the model. For this purpose, physical experiment has been carried out. The experimental setup shown in Fig. 5 was used to carry out the experiment. The 3 Degree-of-Freedom (DOF) motion platform drives the needle into the phantom/animal organ following some pre-designed trajectory. A 6 DOF force/torque (F/T) sensor (Nano17-SI-12-0.12 ® from ATI Industrial Automation, USA) is mounted at the needle end to measure the needle end force. The needle adopted here is a 5-DOF MagTrax Needle Probe (from Traxtal Technology, Canada). It is a 130mm long needle and has a sensor located at the stylet's proximal tip. This needle movement can be observed in real time via an electromagnetic system called Aurora® (from Northern Digital Inc. Canada).

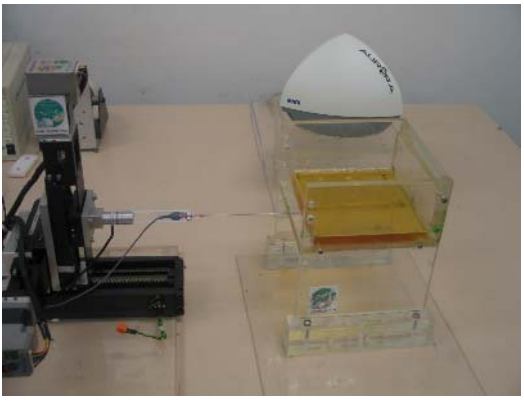


Fig. 5. Experiment setup

Phantoms made of different gelatin/water ratios were firstly adopted to simulate the soft tissue, for it's easier to obtain and the properties are easier to control. The

needle was driven into the phantoms with insertion speeds at 10mm/s, 8mm/s, 6mm/s and 4mm/s, respectively. The needle tip trajectory and the end force were collected during the insertion. The input force data and output tip position data were filled into the estimator first to estimate the parameters. The estimated parameters and input force data were then put into the model for simulation. For the space constraint, only one result is shown here. The simulation result vs. the tracked output is shown in Fig. 6. The big drop-down at the beginning was caused by the sensor noise when the needle was accelerated to reach the set constant velocities. We can see that the proposed model has predicted the needle tip trajectory quite well.

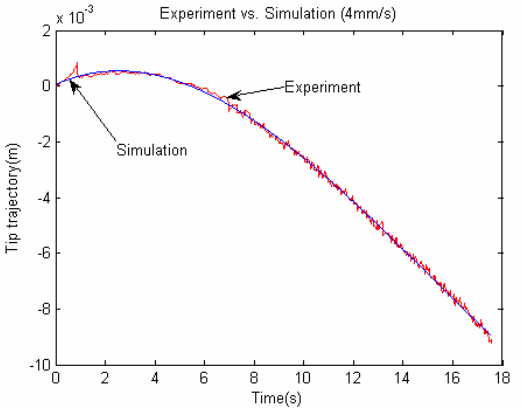


Fig. 6. Experiment vs. Simulation

5 Conclusions and Future Works

In this paper, we proposed an online parameter estimator using EKF for parameter estimation of the needle steering model. Simulation has been carried out first to validate this estimator before used for the steering model. Results have shown that the simulator can estimate the parameters fast and accurately with reasonably set initial covariance matrix. Even though the setting of the initial covariance matrix is rather rule-of-thumb, it has shown that, the magnitudes of 10^5 can always give good convergence rate. Exceeding this magnitude may result in fluctuation and overshoot.

The steering model was then validated though physical experiment in gelatin phantom. The needle end force and needle tip position data were collected and put into the parameter estimator to estimate the parameters. The EKF method makes use of the minimum variance estimation that is optimal for each point in time, not a sum over time. This reversely ascertains that the output of the steering model when using the estimated parameters can generate the output that is the best fit for the measured output.

At current stage, the spring and damper coefficients are assumed to be constant. Considering the inhomogeneous human tissue, in the near future, we will improve them to be depth varying and modify the online parameter estimator accordingly for more robust performance.

Acknowledgments. Nanyang Technological University (Singapore), NMRC 0859/2004 and US National Cancer Institute (NCI) (under grant R01 CA091763) are acknowledged for support.

References

1. Loser, M, Navab, N: A New Robotic System for Visually Controlled Percutaneous Interventions under CT. MICCAI. Lecture notes in computer science, Vol 1935, Springer Verlag, (2000)
2. Stoianovici, D., Cadeddu, J. A., Demaree, R. D., Basile, H. A., etc.: An Efficient Needle Injection Technique and Radiological Guidance Method for Percutaneous Procedures. Lecture Notes in Computer Science, Vol. 1205, Springer Verlag, (1997)
3. Kaiguo, Yan, Wan Sing Ng, Yan Yu, Tarun Podder, Tien-I Liu, CWS Cheng, Keck Voon Ling: Needle Steering Modeling and Analysis using Unconstrained Modal Analysis. BIOROB, Italy, (2006)
4. D. Terzopoulos and K. Waters: Analysis and Synthesis of Facial Image Sequences using Physical and Anatomical Models. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 15, (1993)
5. F. Boux de Casson and C. Laugier: Modeling the Dynamics of a Human Liver for a Minimally Invasive Surgery Simulator. MICCAI, Lecture Notes in Computer Science, Vol.1679, (1999)
6. Webster R: Elastically Deformable 3D Organs for Haptic Surgical Simulation. Proceedings of Medicine Meets Virtual Reality, Newport Beach, (2002)
7. P. F. Neumann, L. L. Sadler, and J. G. M.D: Virtual Reality Vitrectomy Simulator. MICCAI, Lecture Notes in Computer Science, Vol.1496, (1998)
8. Carlos Canudas de Wit, Bruno Siciliano and Georges Bastin: Theory of Robot Control, Springer, New York, (1996)
9. Martin Morf and Thomas Kailath: Square-Root Algorithms for Least-Squares Estimation. Transaction on Automatic Control, Vol. AC-20, (1975)
10. Mi Lu and Xiangzhen Qiao: Parallel Computation of the Modified Extended Kalman Filter. International Journal of Computer Math, Vol. 45, (1992)
11. Satoko Abiko, Kazuya Yoshida: On-line Parameter Identification of a Payload Handeled by Flexible Based Manimulator. Proceedings of 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, Japan, (2004)
12. Tien Li Chia, Po-Chuan Chow, Howard Jay Chizeck: Recursive Parameter Identification of Constrained Systems: An Application to Electrically Stimulated Muscle. IEEE Transactions on Biomedical Engineering, Vol. 38, (1991)
13. Akihiko Kumagai, Tien-I Liu, and Paul Holzian: Control of Shape Memory Alloy Actuators with A Neuro-Fuzzy Feedforward Model Element, Journal of Intelligent Manufacturing, Vol. 17, 45-56, (2006)
14. Paul Zarchan and Howard Musoff: Fundamentals of Kalman Filtering, A Practical Approach, the American Institute of Aeronautics and Astronautics, Inc.1801 Alexander Bell Drive, Reston, Virginia 20191-4344,Vol. 190, (2000)

Realistic Simulated MRI and SPECT Databases
Application to SPECT/MRI Registration Evaluation

Berengere Aubert-Broche[1], Christophe Grova[1], Anthonin Reilhac[1,2], Alan C. Evans[1], and D. Louis Collins[1]

[1] Montreal Neurological Institute, McGill University, Montreal, Canada
[2] CERMEP, Lyon, France

Abstract. This paper describes the construction of simulated SPECT and MRI databases that account for realistic anatomical and functional variability. The data is used as a gold-standard to evaluate four SPECT/MRI similarity-based registration methods.

Simulation realism was accounted for using accurate physical models of data generation and acquisition. MRI and SPECT simulations were generated from three subjects to take into account inter-subject anatomical variability. Functional SPECT data were computed from six functional models of brain perfusion. Previous models of normal perfusion and ictal perfusion observed in Mesial Temporal Lobe Epilepsy (MTLE) were considered to generate functional variability. We studied the impact noise and intensity non-uniformity in MRI simulations and SPECT scatter correction may have on registration accuracy.

We quantified the amount of registration error caused by anatomical and functional variability. Registration involving ictal data was less accurate than registration involving normal data. MR intensity non-uniformity was the main factor decreasing registration accuracy. The proposed simulated database is promising to evaluate many functional neuroimaging methods, involving MRI and SPECT data.

1 Introduction

Digital simulations of imaging modalities provide a way of generating data where a gold-standard is available. They become widely used as evaluation datasets to characterize and optimize the performance of image processing methods. Simulators accurately modelling physical properties of MRI, PET or SPECT data acquisition have been proposed and validated [1, 2, 3]. Realistic simulations are usually obtained from high-resolution human brain data, defining a spatial model of brain anatomy. Such models were based on a single brain [4],[5]. Recently, to take into-account inter-subject anatomical variability for MRI simulations, we built anatomical models from multiple T1, T2 and PD-weighted MRI acquisitions of 20 normal adults (submitted work). Reilhac *et al.* [3] used anatomical MRI from 17 normal adults to simulate PET data.

For realistic simulations of functional data such as SPECT and PET, anatomical models are required to define the attenuation map describing the attenuation properties of head tissues and the activity map representing the 3D spatial dis-

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 330–337, 2006.

tribution of the radiotracer. Generating a realistic activity map usually requires measurements on real data, as for instance averaged normal perfusion and averaged ictal perfusion models [6]. Functional variability may be characterized by exploring the structure of covarying areas among functional data issued from homogeneous populations. Principal component analysis [7, 8] and correspondence analysis [9] have proved their ability to characterize normal perfusion and ictal perfusion in patients with MTLE.

In this paper, we propose to adapt anatomical models previously developed to simulate SPECT data. To simulate functional variability, three maps mimicking normal perfusion and three maps mimicking ictal perfusion in MTLE were generated from the main modes of variation of the model proposed in [9]. For each subject and each functional model, MR simulated data were generated with different noise and intensity non-uniformity levels. Simulated SPECT data were reconstructed with and without scatter post-correction. These SPECT and MRI simulated data-sets were used to evaluate four similarity-based registration methods : mutual information (MI), normalized mutual information (NMI), L1 and L2 norm-based correlation ratios (CR).

2 Methods

2.1 MR Simulations

Anatomical Models. The method proposed in [10] was used to create anatomical models from 20 healthy subjects. For each subject, we acquired four T1, T2 and PD-weighted scans. All images were corrected for intensity non-uniformity [11], registered [12] and resampled in the average brain space [13], intensity normalized and averaged to create averaged T1, T2 and PD-weighted volumes for each subject. A fuzzy bayesian classification was used to classify voxel intensities from T1, T2 and PD averaged volumes into grey matter, white matter, cerebro-spinal fluid and fat. Automatically generated mask volumes were employed to separate brain from non-brain structures and to create the following classes of the anatomical model: grey matter, white matter, cerebro-spinal fluid, skull, bone marrow, dura, fat, muscles and skin.

Simulation. Each class of the anatomical model was described by its nuclear magnetic resonance (NMR) relaxation properties. NMR signal intensities were computed for all classes by feeling these NMR parameters into a discrete-event Bloch equation in the MR simulator [1]. The simulator could model noise and RF field inhomogeneity. T1-weighted MRI data were simulated at two levels of noise, i.e., 3% (named 'mri-pn3') and 9% (named 'mri-pn9') standard deviation. A high level of intensity non-uniformity (INU) (60%, named 'mri-rf60') was also simulated from data generated at 3% of noise. Scan parameters were similar to those used during a clinical acquisition (3D spoiled gradient echo (GRE) sequence, $TR = 22ms, TE = 9.2ms, \alpha = 30^o$, 1 mm isotropic voxel size).

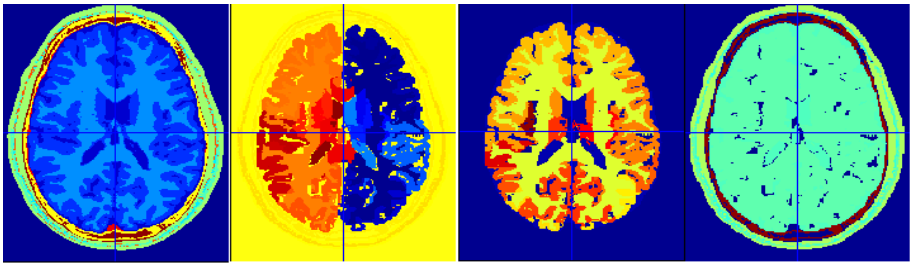
2.2 SPECT Simulations

Anatomical Models. For each subject, the grey matter of the MRI anatomical models was subdivided to define the SPECT anatomical models. A digital atlas was first mapped onto the grey matter class, using a non linear registration algorithm to automatically extract the different grey matter regions [14]. These grey matter regions were modified as the anatomical regions usually considered by epilepsy specialists divide the temporal and frontal lobes into their posterior, lateral, mesial and polar areas [6].

Attenuation and Activity Maps. The attenuation map was obtained by assigning an attenuation coefficient μ at the 140 keV energy emission of ^{99m}Tc to each class of the anatomical model. Seven tissue types were considered : connective tissue, water, brain, bone, muscle, fat and blood. The model proposed in [9] was used to generate realistic activity maps accounting for functional variability. 27 SPECT data from healthy subjects and 10 ictal SPECT from patients with MTLE were first spatially and intensity normalized into an average brain space. The digital atlas described above was then used to perform Volume of Interest (VOI)-based measurements on each SPECT data. The structure of covariance among these VOI measurements over all SPECT data was measured using correspondence analysis. Five principal components accounting for 67.5% of total variance were selected to model normal perfusion. Two principal components accounting for 69.2% of total variance were considered for ictal perfusion. Three healthy subjects and three patients with MTLE were first randomly selected. Six activity maps mimicking normal perfusion (N1, N2, N3) and ictal perfusion in MTLE (I1, I2, I3) were then generated by projecting the VOI measurements of each subject on the corresponding selected principal components. This projection aims at removing measurement noise in the data in order to account only for the main modes of perfusion variability.

Simulation and Reconstruction. Using attenuation and activity maps, Monte Carlo simulations were performed using the SimSET[1] simulator [15]. Projections were simulated using a 20% energy window centred on 140 keV (126-154 keV) and a Compton window (111-125 keV). Attenuation and Compton scatter were modelled through Monte Carlo methods. The collimator and gamma camera responses were modelled through analytical models. To assess the impact of scatter correction on registration accuracy, Jaszczak scatter correction was applied by subtracting the simulated projections of the Compton window from the simulated projections of the 20 % energy window centred on 140 keV. All projections were reconstructed by filtered backprojection using a ramp filter (Nyquist frequency cutoff). The reconstructed data (4.51 mm isotropic voxel size) were post-filtered with a 3D Gaussian filtering (FWHM = 8mm).

[1] Simulation of Emission Tomography (SimSET) package, *http : //depts.washington.edu/ simset/html/simset_main.html*



(A) (B) (C) (D)

Fig. 1. For the first subject,(A) anatomical model used for MRI simulation, (B) grey matter subdivided-model used for SPECT simulation, (C) an activity map modelling normal perfusion and (D) an attenuation map obtained from this model

2.3 SPECT/MRI Registration Evaluation

Registration Methods. These simulated data were used to evaluate four statistical similarity-based SPECT/MRI registration methods: mutual information (MI), normalized mutual information (NMI), L1 and L2 norm-based correlation ratios (CRL1 and CRL2). Each registration method was implemented as proposed by [16].

Evaluation. We studied the impact anatomical and functional variabilities generated by the proposed simulated data have on registration performance. For each subject and for each functional model, we computed registrations between three simulated MRI data (two with different noise levels and one with INU) and two SPECT reconstructions (with and without photon scatter correction). Table 1 summarize the different simulations contexts. For each subject, the three simulated MR volumes are registered to twelve SPECT volumes using four different similarity measures.

Since the simulated SPECT data was perfectly aligned with simulated MRI data by construction, we applied a known geometrical transformation to the simulated MRI data to generate mis-aligned data. For each simulation context, 20 known transformations were generated by randomly sampling a six parameters vector using a Gaussian distribution (mean = 0, standard deviation = 10 mm or ^{o}). Comparing the known geometrical transformation to the resulting computed transformation, local target registration errors (TRE) were estimated on 200 points uniformly distributed within the brain of each subject. The root mean square value (RMS) of the local errors distribution was estimated to characterize the spatial distribution of TRE within the brain.

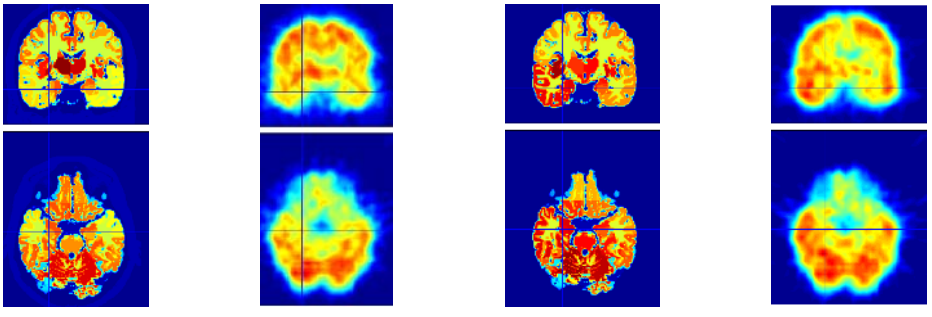
3 Results

3.1 MRI and SPECT Simulations

For the first subject, Fig. 1 shows transverse slices of the anatomical model used for MRI simulation, the grey matter subdivided-model used for SPECT simula-

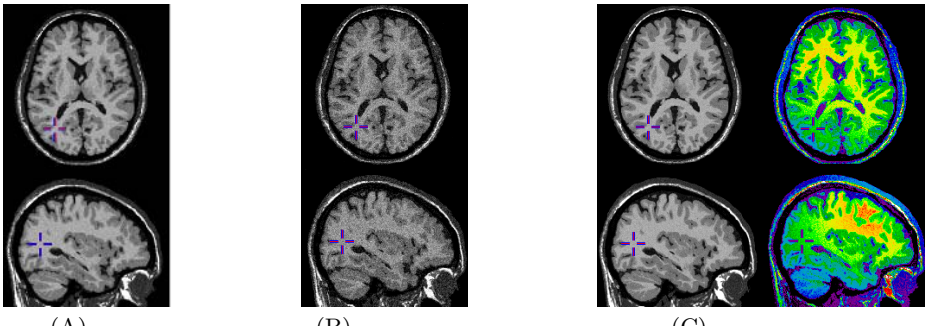
Table 1. Simulation contexts explored by SPECT/MRI registration evaluation

Anatomical model	Functional model	MRI data quality	SPECT data quality	Registration methods
3 subjects	- 3 normal models - 3 ictal models	- 3% noise - 9% noise - 3% noise and INU	- scatter correction - no scatter correction	- CRL1 - CRL2 - MI - NMI



(A) (B)

Fig. 2. For the first subject, activity maps and simulated spect (A) with normal perfusion N1 and (B) with ictal perfusion I1



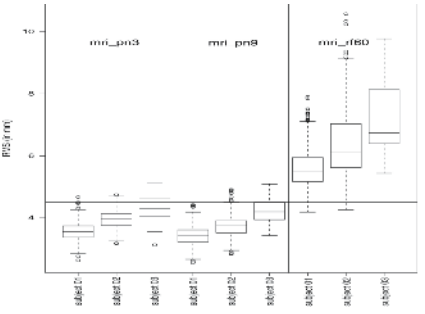
(A) (B) (C)

Fig. 3. For the first subject, simulated MRI (A) with 3% noise, (B) with 9% noise and (C) with 3% noise and 60% intensity non-uniformity

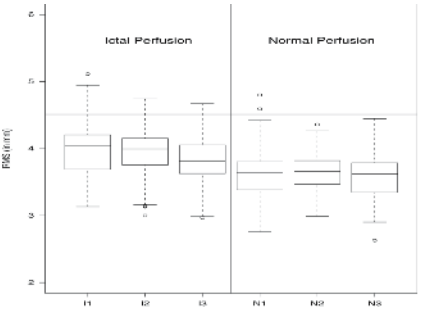
tion, a normal activity map and an attenuation map obtained from this model. To illustrate functional variability, two activity maps and resulting simulated SPECT are shown Fig. 2. Fig. 3 shows MRI simulated data of the same subject for different noise levels and INU.

3.2 Registration Evaluation

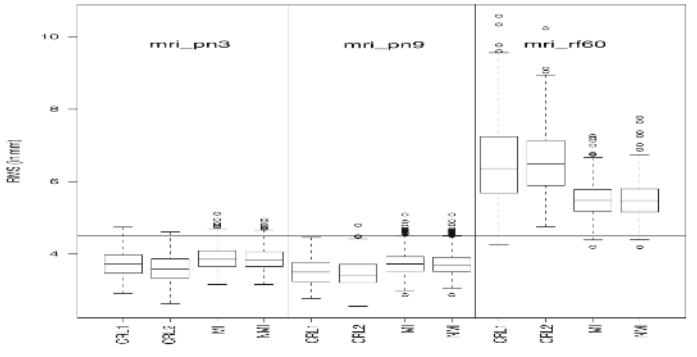
Distribution of registration RMS errors (in mm) are summarized using boxplot distributions in Fig. 4. Whereas increasing levels of noise in MRI data had slight



(a) Effect of anatomical variability (3 subjects)



(b) Effect of functional variability (6 functional models: ictal perfusion I1, I2 and I3 and normal perfusion N1, N2 and N3 with mri-pn3)



(c) Effect of the registration method

Fig. 4. Distributions of SPECT/MRI registration errors (RMS in mm) for different MRI data "quality": 3% noise (mri-pn3), 9% noise (mri-pn9) and 3% noise with intensity non-uniformity (mri-rf60). Results obtained for all parameters omitted in the captions were pooled together in the boxplot representations. Horizontal line indicates SPECT voxel size at 4.51 mm

impact on registration accuracy, generating high intensity non-uniformity level on MRI data (60%) clearly decreased registration accuracy (Fig. 4(a) and (c)). We showed also an impact of anatomical variability on registration performance (Fig. 4(a)). Analysis of variance confirms a significant effect of the subjects and MRI data quality on registration accuracy (F=3834, R^2_{adjust}=0.82, p <0.001). We found no effect of scatter post-correction of SPECT data on registration accuracy (results not shown).

We observed a significant effect of the clinical context (ictal vs normal perfusion) on registration accuracy (analysis of variance, F=232.5, R^2_{adjust}=0.21, p <0.001). Registrations involving normal SPECT simulations were significantly

more accurate than registrations involving ictal data (Fig. 4(b)). Although small variations due to functional variability were observed within each perfusion group (ictal or normal)(Fig. 4(b)), main trends concerning the loss of registration accuracy in pathological condition were still present. In presence of INU, MI and NMI based registrations were slightly more accurate than CRL1 and CRL2 methods (Fig. 4(c)). The opposite situation was observed when considering MRI simulations without INU.

4 Discussion and Conclusion

The simulation database computed for this study enables one to take into account both anatomical variability, using anatomical models from three subjects, and functional variability, considering six functional models (ictal and normal). Using fully controlled simulated data, the impact of many data quality parameters could be studied, such as noise and INU levels in MRI data, or SPECT data scatter correction. Accounting for anatomical and functional variability in simulated databases is essential for validation of statistical analysis methods, as for instance epileptogenic focus detection methods or inter-subject group analysis methods. We plan to make these images publicly available on the internet for comparison of different methods.

Assessing the realism of the simulations is an important and difficult issue. All simulated data rely on techniques reproducing accurately the physics of data generation and acquisition [1, 2]. To evaluate the realism of simulated data, previous studies reported comparison between mean intensities measurements between real and simulated SPECT data [6]. From the 20 subjects acquired for this project, voxel-based group comparison between simulated and real T1-weighted data confirmed the realism of MRI simulations (submitted work). We plan to use a similar approach to assess the realism of our simulated SPECT data. Note that the realism of the method used to characterize perfusion variability should also be assessed, using a leaving-one-out evaluation procedure for instance [7].

The SPECT/MRI registrations evaluated in this study illustrate an important application of SPECT and MRI simulated databases. We quantified the amount of registration error variability caused by simulated anatomical and functional variability. Our results were in agreement with [16], showing that registration of pathological data was less accurate than registration of normal data, due to an increase in intensity dissimilarity between data. The major result was the important decrease of registration accuracy due to a high intensity non-uniformity on MRI data, which has never been studied before in the context of SPECT/MRI registration. However, the effect of MRI intensity non-uniformity was demonstrated for T1-T2 NMI registration [17]. The proposed framework is ideal to quantify the impact of several INU levels on registration accuracy or to quantify if correction methods [11] may improve SPECT/MRI registration accuracy.

We are currently working on the extension of this simulated database. We plan to use the 20 anatomical models already available, as well as more perfusion models, using random sampling of activity values within the main modes of variation.

References

1. R. Kwan, A.C. Evans, and G.B. Pike. MRI simulation-based evaluation of image-processing and classification methods. *IEEE Trans Med Imaging*, 18(11):1085–1097, 1999.
2. I. Buvat and I. Castiglioni. Monte carlo simulations in SPET and PET. *Quarterly J. Nucl. Med.*, 46:48–59, 2002.
3. A. Reilhac, G. Batan, C. Michel *et al.* PET-SORTEO: validation and development of database of simulated PET volumes. *IEEE Trans Nucl Sci.*, 52(5):1321–1328, 2005.
4. IG Zubal, CR Harrell, EO Smith *et al.* Computerized three-dimensional segmented human anatomy. *Med Phys*, 21(2):299–302, 1994.
5. D.L. Collins, A.P. Zijdenbos, V. Kollokian *et al.* Design and Construction of a Realistic Digital Brain Phantom. *IEEE Trans Med Imaging*, 17(3):463–468, June 1998.
6. C. Grova, P. Jannin, A. Biraben *et al.* A methodology for generating normal and pathological brain perfusion spect images for evaluation of mri/spect fusion methods: application in epilepsy. *Phys Med Biol*, 48(24):4023–4043, 2003.
7. T. Vik, F. Heitz, I. Namer, and J-P. Armspach. On the modeling, construction, and evaluation of a probabilistic atlas of brain perfusion. *Neuroimage*, 24(4):1088–1098, 2005.
8. BJ Weder, K Schindler, TJ Loher *et al.* Brain areas involved in medial temporal lobe seizures: A principal component analysis of ictal spect data. *Human Brain Mapp*, 2005.
9. C. Grova, P. Jannin, I. Buvat *et al.* From anatomic standardization analysis of perfusion spect data to perfusion pattern modeling: evidence of functional networks in healthy subjects and temporal lobe epilepsy patients. *Acad Radiol*, 12(5):554–565, 2005.
10. B Aubert-Broche, DL Collins, and AC Evans. A new improved version of the realistic digital brain phantom. *Neuroimage*, 2006. in press.
11. J.G. Sled, A.P. Zijdenbos, and A. Evans. A Nonparametric Method for Automatic Correction of Intensity Nonuniformity in MRI Data. *IEEE Trans Med Imaging*, 17(1):87–97, February 1998.
12. DL Collins, P Neelin, TM Peters, and AC Evans. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J Comput Assist Tomogr*, 18(2):192–205, Mar-Apr 1994.
13. A.C. Evans, D.L. Collins, S.R. Mills *et al.* 3D statistical neuroanatomical models from 305 MRI volumes. In *IEEE Nuclear Science Symposium and Medical Imaging Conference*, pages 1813–1817, San Francisco, USA, October 1993.
14. DL Collins, AP Zijdenbos, WFC Barre, and Evans AC. ANIMAL+INSECT: Improved cortical structure segmentation. In *Proc. of the Annual Symposium on IPMI*, pages 210–223, 1999.
15. R.L. Harrison, S.D. Vannoy, D.R. Haynor *et al.* Preliminary experience with the photon history generator module of a public-domain simulation system for emission tomography. *IEEE Nucl Sci. Symposium*, 2:1154–1158, 1993.
16. C Grova, P Jannin, I Buvat *et al.* Evaluation of registration of ictal SPECT/MRI data using statistical similarity methods. In *MICCAI (1)*, pages 687–695, Saint-Malo, France, September 2004.
17. ZF Knops, JB Maintz, MA Viergever, and JP Pluim. Normalized mutual information based registration using k-means clustering and shading correction. *Med Image Anal.*, 2005.

Extrapolating Tumor Invasion Margins for Physiologically Determined Radiotherapy Regions

Ender Konukoğlu[1], Olivier Clatz[1], Pierre-Yves Bondiau[2], Hervé Delingette[1], and Nicholas Ayache[1]

[1] Asclepios Research Project, INRIA Sophia Antipolis, France
ender.konukoglu@sophia.inria.fr
[2] Centre Antoine Lacassagne, Nice, France

Abstract. In radiotherapy, the constant margin taken around the visible tumor is a very coarse approximation of the invasion margin of cancerous cells. In this article, a new formulation to estimate the invasion margin of a tumor by extrapolating low tumor densities in magnetic resonance images (MRIs) is proposed. The current imaging techniques are able to show parts of the tumor where cancerous cells are dense enough. However, tissue parts containing small number of tumor cells are not enhanced in images. We propose a way to estimate these parts using the tumor mass visible in the image. Our formulation is based on the Fisher-Kolmogorov Equation that is been widely used to model the growth of brain tumors. As a proof of concept, we show some promising preliminary results, which demonstrate the feasibility of the approach.

1 Introduction

Glial based tumors form the major class of tumors in the central nervous system. Among this class, high grade ones are incurable despite state-of-the-art therapies, and patients have a median survival rate of approximately 1 year, [1]. The diffusive nature of these tumors inhibits the apparent boundary between cancerous and healthy regions, decreasing effects of conventional therapy, like surgical resection and radiotherapy. The current imaging techniques are able to detect only a part of the total tumor, [2]. In the literature there are several suggestions for the lowest detection limit of CT images in terms number of tumor cells per unit volume or area, [2, 3, 4]. Based on these suggestions and medical advice, in this work we are using a relative value: 5% of the maximum number of tumor cells brain paranchyma can hold, as this detection limit. Although up to our knowledge, there is no such a limit suggested for MRI it is the common practice to assume the same limit for T2 weighted images. In Figure 1 we illustrate the visible tumor profile in the image (the solid line curve) and the total tumor profile as given in theory. The usual clinical practice is to surgically remove the visible part and try to kill tumor cells invisible in the images, using radiotherapy and/or chemotherapy. The conventional radiotherapy treatment

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 338–346, 2006.

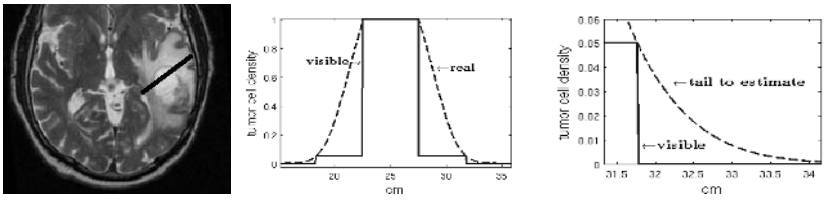


Fig. 1. Left: T2-weighted MR image with the tumor, Middle: Visible tumor density profile for the black line of the left image is given in solid curve and the theoretical tumor density profile shown by the dashed curve, Right: Regions with less tumor density than 5% are not visible in the image. The aim of this paper is to estimate cell density distribution in these regions (dashed curve).

uses a geometric approach and applies radiation in a constant margin of about 2 cm, around the surgical site. This approach does not take into account different speed of tumor diffusion for the grey and the white matter, [5]. Using the same margin for both matters may kill healthy grey tissue while not reaching the extent of tumor cells in the white matter, leaving alive cancerous cells on the fibers. These cells would cause reccurence of the tumor in time. By constructing physiologically determined radiotherapy margin, which takes into account tumor growth dynamics and the properties of the tissue, the therapy can be more effective.

Mathematical modeling of tumor growth dynamics gives us a better insight on the physiology of this process and it could be used to improve the treatment planning (surgery, radiotherapy or chemotherapy). There is a lot of work on modelling the growth of high grade gliomas. Approaches taken to solve this problem can be coarsely classified into two groups, microscopic and macroscopic ones. Microscopic ones try to describe cell divisions and invasion by means of interactions between tumor cells and their surrounding tissue, [6, 7]. On the other hand, macroscopic models describe the evolution of local tumor cell densities and try to capture the dynamics by general equations and with fewer parameters. Moreover, such models may more easily be adapted to act on real medical images. Most of the macroscopic models, [1, 3, 4], are based on the reaction-diffusion formalism introduced by Murray in [8, 1]. This formalism uses the general type of PDEs called the reaction-diffusion to model the tumor growth dynamics.

We are proposing a method to estimate the distribution of tumor cells for low density regions using a single image with a set of estimated parameters. Our formulation constructs an approximation to the tails of the tumor distribution (invasion margins) which are not visible in images. In addition, this formulation creates appropriate initial conditions for tumor growth models, which can be used for simulating tumor growth and predicting the distribution of the tumor in a future time. Using an initial condition with well estimated tails will increase ability of these models to predict tails of tumor as well as its whole distribution.

2 Method

2.1 Fisher-Kolmogorov Equation

In this article, we assume that tumor growth dynamics obey a reaction-diffusion Equation similar to the ones proposed in [3, 4], in which diffusion of tumor cells is correlated with the diffusion of water, hence with the DT-MRI. The specific Equation we will use is the Fisher-Kolmogorov (F-KPP) Equation given as:

$$\frac{\partial u}{\partial t} = \nabla \cdot (D(\mathbf{x})\nabla u) + \rho u(1-u) \tag{1}$$

$$D\nabla u \cdot \overrightarrow{n}_{\Sigma} = 0 \tag{2}$$

where u is the normalized tumor cell density between $[0, 1]$ (normalized cell densities will be used throughout this article), D represents the diffusion tensor of tumor cells and ρ corresponds to the proliferation rate of tumor cells. In Equation 1, the first term of the right hand-side describes the diffusion of tumor cells towards the tissue, which is governed by D. The second term, $\rho u(1-u)$, is the so called reaction term, and it describes proliferation of tumor cells as a logistic growth. There are different types of reaction terms that can be used like the gompertz growth model, $\rho u \ln(1/u)$, or exponential growth model, ρu. Equation 2 represents the no-flux boundary condition, where Σ is the boundary of the brain (ventricles and the skull) and $\overrightarrow{n}_{\Sigma}$ is the normal to the boundary. This condition describes the fact that tumor cells do not diffuse through ventricles and through the skull.

The simple version of this Equation in 1D with constant coefficients and no boundary conditions was first proposed by Fisher, [9], to model the spread of a favoured gene in a population:

$$\frac{\partial u}{\partial t} = d\frac{\partial^2 u}{\partial x^2} + \rho u(1-u) \tag{3}$$

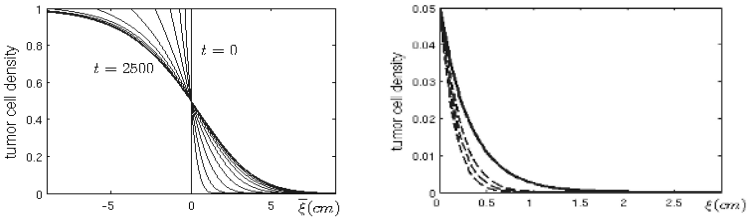
where d is the diffusion coefficient. The travelling wave solutions of Equation 3 have been studied throughly by Kolmogorov. Later on Aronson et al. have extended this analysis to the infinite cylinder and studied the travelling plane solutions in [10].

Here we wish to summarize some relevant results in these works. At large times, Equation 3 admits a travelling wave solution in the infinite cylinder. In other words, when the change of u is non-zero in only one direction, $\mathbf{n}$, for very large times the solution can be given in the form:

$$u(\mathbf{x}, t) = u(\mathbf{n} \cdot \mathbf{x} - ct) = u(\overline{\xi}) \quad \text{as} \quad t \to \infty \tag{4}$$

where c corresponds to the speed of the front and $\overline{\xi} = (\mathbf{x} \cdot \mathbf{n} - ct)$ is the moving frame. The speed of the front is a constant, which depends on coefficients D and ρ, and also on the tails of the initial condition $u(\mathbf{x}, 0)$. When the initial condition has a compact support the speed of the travelling wave can be given as, [10]:

$$c = 2\sqrt{\rho \mathbf{n} \cdot (D\mathbf{n})} \tag{5}$$



(a) (b)

Fig. 2. (a) Shape of the wavefront is plotted for different time values (0-2500 days) in the moving frame. The profile converges in time. (b) The approximated tail using Equation 8 (solid curve) and actual tails of the wave front at t = 60,90,125 days (from left to right). As time increases the solid curve approximates the actual tail better.

The planar initial condition with compact support converges to a travelling plane with speed c in time. As the speed of the travelling plane converges to c the shape of it also converges to a constant shape, which does not have an analytical form. Figure 2(a) illustrates the convergence of the front shape both for whole front and for the low density parts.

2.2 Tumor Tail Extrapolation

Our aim in this article is to extrapolate the low tumor density regions, invisible in the image, from the visible part of the tumor, at a given time instant. We are going to use the fact that the front shape is constant for the travelling plane solution for large times to extrapolate the low density regions.

Since we are trying to extrapolate the tails (low values of u) we propose to linearize the nonlinear reaction term in Equation 1 around $u = 0$ and assume $1 - u$ can be well approximated by 1:

$$\frac{\partial u}{\partial t} = D\Delta u + \rho u \tag{6}$$

This Equation also admits a travelling wave solution in the infinite cylinder, with the speed c given in Equation 5. The travelling wave solution for this Equation has an analytical solution, and for the low tumor density parts the solution can be given as:

$$u(\mathbf{x}, t) = u(\overline{\xi}) \cong Ae^{-\lambda\overline{\xi}} \quad \text{for} \quad \lambda = -\sqrt{\rho/(\mathbf{n} \cdot (D\mathbf{n}))} \tag{7}$$

where A is an integration constant.

The value of $\overline{\xi}$ for a point corresponds to the distance of it from the inflection point of the front shape, which is at $u = 0.5$ (see Figure 2(a)). However, the only information we get from images is the location of the last visible iso-density contour u_0, which we assume to be equal to 0.05, in terms of normalized tumor densities, see Figure 1. Applying the solution given in Equation 7 for the value u_0, we find $A = u_0 e^{\lambda\overline{\xi}_0}$ where $\overline{\xi}_0 = \{\overline{\xi} \mid u(\overline{\xi}) = u_0\}$. Placing this in Equation 7:

$$u(\mathbf{x},t) \cong u_0 e^{-\lambda(\overline{\xi}-\overline{\xi}_0)} = u_0 e^{-\lambda\xi} \tag{8}$$

where $\xi = \overline{\xi} - \overline{\xi}_0$, which is the distance of a point from the iso-density surface $u = u_0$. Unlike $\overline{\xi}$, ξ is observable in the images since the u_0 iso-density contour is visible. In Figure 2(b) we plot the tail approximation given by Equation 8 as a function of ξ together with the front shape of the travelling plane taken at different time instants.

This approximation is only valid when the diffusion tensor is constant over the whole domain and the motion is only in one direction, which is not the case. Moreover when the tumor front is curved, its motion is not in one direction and the solution cannot be given in terms of travelling wave. However, to compute the tumor cell densities at tails in MR images we make the following hypothesis: D is constant and the motion of the front is only in one direction within a voxel. Based on this assumption we construct the solution given in Equation 7 for each voxel. The direction of motion and the initial value for each voxel are defined by adjacent voxels. The 1-dimensional illustration of this idea is given in Figure 3. Equation 7 is the integral solution of $\partial u/\partial \mathbf{n} = (\lambda\xi)u$, since ξ direction is normal

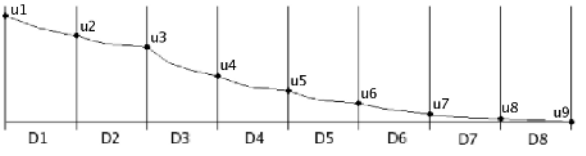


Fig. 3. The formulation constructs the solution at grid point $p+1$ using the D_p as the diffusion tensor and u_p as the initial value

to the front. Replacing $\mathbf{n}$ with $\nabla u / \mid \nabla u \mid$ in this Equation, we obtain the following static Hamilton-Jacobi Equation that constructs the solution given by Equation 7 at each voxel as shown in Figure 3.

$$\frac{\sqrt{\nabla u \cdot (D\nabla u)}}{\sqrt{\rho} u} = 1,\ u(\Gamma) = u_0 \tag{9}$$

where Γ is the last visible iso-density contour of the tumor in the image. This Equation has two solutions at each point, one with increasing u and one with decreasing u. Since the F-KPP Equation tells us that as we move away from the visible contour, the values of u will decrease, for all points we choose the decreasing solution. Using Equation 9 we start from Γ and move outwards as we find u values for each point. We also have to include the no-flux boundary condition given in Equation 1 in this approximation. Let Ω be the whole image domain and Ω_B be ventricles and the skull with its exterior. We enforce the no-flux boundary condition by setting the domain of Equation 9 as $\Omega_s = \Omega / \overline{\Omega_B}$. In which $\overline{\Omega_B}$ is the closure of Ω_B.

2.3 Numerical Implementation

There are several numerical techniques proposed to solve static H-J Equations, amongst which sweeping methods, such as fast marching, are the most popular. However, we do not use the sweeping methods because they construct the integral solution at the center of each voxel, for which the accuracy is low when the resolution of the image is low. In order to solve Equation 9 numerically we embed the static H-J Equation into a dynamic one using the method proposed by Osher in [11]. This methodology allows us to make use of the level set idea and obtain subvoxel accuracy. This is crucial since the resolution of MR images is not high ($1mm \times 1mm \times 4mm$). The embedding is done by placing each iso-density contour of u as the zero-level set of a higher dimensional function v. Namely, the embedding is such that, $u(\xi) = \tau$ iso-density contour corresponds to the $v(\xi, \tau) = 0$. Using this we transform Equation 9 into Equation 10.

$$\frac{\partial v(\xi,\tau)}{\partial \tau} = \frac{\sqrt{\nabla v \cdot (D \nabla v)}}{\sqrt{\rho}\tau} \quad \text{for } \tau \in (0, u_0] \tag{10}$$

where τ decreases at every iteration so that we find lower tumor density regions. As the zero level-set of v moves over the domain of u, Ω_s, for different τ we construct the u value at each point as $u(\xi) = \tau$ for $\xi \mid v(\xi, \tau) = 0$. We have used the method proposed by Bryson et al. in [12] to solve this dynamic H-J Equation numerically in 3D, due to its ease of implementation and accuracy properties. Using an adaptive τ step as: $\Delta\tau = \tau/30$ the algorithm takes around 30 seconds to compute the tails in the whole brain until $u = 0.0005$.

As of implementing Equation 1 in 3D, we have used finite differences with a semi-implicit solver (pre-conditioned conjugate gradient). The boundary conditions for the diffusion part given in Equation 2 were included in the numerical scheme by using the flux method proposed by McCorquodale et al. in [13].

3 Results

We have run our experiments by artifically placing tumors on T2-weighted images taken from healthy subjects. We assume Equation 1 models the tumor growth accurately, so we compare our results using a synthetic tumor grown with this model, for which we know the whole distribution. Giese et al. in [5] have shown *in-vitro* that tumor cells move faster on myelin sheath. We use this information and a correlation between water and tumor diffusion to construct our tumor diffusion tensor, D, from DT-MR images of subjects as done in [3, 4]. We use $D_{grey} = d_0\mathbf{I}$, where $\mathbf{I}$ is the identity matrix, creating isotropic diffusion tensor D_{grey} for the grey matter. $D_{white} = d_0 \alpha D_{water}$ is the diffusion tensor for tumor cells in the white matter, which has the same orientation as the normalized water diffusion D_{water} at that point. α describes the speed difference of tumor cells in the white and the grey matter. In our experiments we have taken α so that the largest diffusion in the white matter would be equal to 50 times

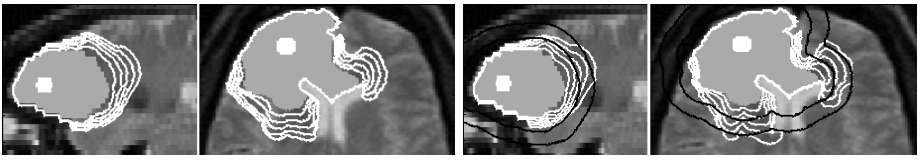


Fig. 4. Left: Sagital and axial views of the T2-weighted image for the synthetic tumor including its iso-density contours from 0.05 to 0.003. Right: Sagital and axial views of the same T2-weighted image with estimated iso-density contours based on our method. Black contours correspond to constant margin radiotherapy regions for 1cm and 2cm.

the isotropic diffusion in the grey matter. The two parameters for the formulation, d_0 and ρ were set using the values proposed in [1] as $0.0013 cm^2/day$ and $0.012 cells/day$ respectively.

This experiment shows the ability of the proposed model in constructing low tumor density regions from a single image. For this, we first grew a synthetic tumor for 180 days. Then we applied the 5% detection threshold and created the T2-weighted image with the synthetic tumor. In Figure 4 on the left we show the saggital and the axial slices of the T2-weighted image with some iso-density contours below the detection threshold. In the same figure on the right we show same images with corresponding iso-density contours extrapolated by the H-J formulation. In both sagital and axial views, we see that the extrapolated tumor tails using H-J formulation are in very good agreement with tails of the synthetic tumor. We can conclude that in case of a synthetic tumor grown by the F-KPP model, given an image and a set of estimated parameters (D and ρ), the proposed formulation is successful in extrapolating the low tumor density parts, which are undetectable by current imaging techniques. We also plotted the conventional radiotherapy regions (in black) on the image with extrapolated tumor tails to show the difference between the constant margin approach and the estimated invasion regions. Observe that by using physiologically determined radiotherapy margin instead of a constant one, more tumor cells can be killed while less healthy tissue is harmed.

4 Discussion

We have proposed a formulation to estimate low tumor density regions in a given MR image, which are undetectable by current imaging techniques, provided a set of parameters $\{D, \rho\}$ either manually adjusted or estimated by the radiotherapist. Our formulation is based on the F-KPP growth model, which is been widely used for simulating tumor growth in medical images. As a proof of concept we have demonstrated the ability of this formulation in constructing tumor tails by comparing it with a synthetic tumor. Based on these results, we conclude that our formulation is successful in using the underlying tissue structure and the dependence of low density regions to the visible part, for extrapolating undetectable parts of the tumor. These promising preliminary results demonstrate the feasibility of this approach for real patient images.

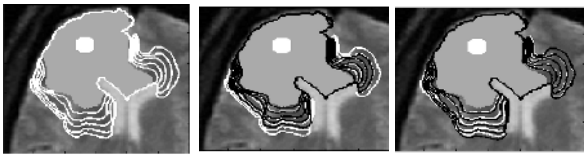


Fig. 5. Left: Undetectable tails of the synthetic tumor are plotted (white). Middle: Contours predicted by F-KPP model (black) using a simple step-edge initialization with actual contours (white). Right: Contours predicted using our tumor tail estimation as an initial condition with actual contours (white).

This formulation constructs an estimate for the undetectable part of the tumor in images, showing the invasion margin, which can be used to determine radiotherapy regions based on tumor growth dynamics. It also provides medical doctors a tool to observe the effect of parameters, like speed of the tumor front and proliferation rate, on undetectable parts of the tumor. Moreover, it constructs an appropriate initial condition for tumor growth models, which can be used for predicting the future distribution of the tumor. Since the initial condition is the key to predict future tumor densities and tail distributions influence these values, using an initial condition with good tail estimates will increase the accuracy of prediction. In Figure 5, we demonstrate the ability of our formulation in creating appropriate initial conditions for the prediction of future tumor cell density. Indeed, we compare two predictions at day 180 (in black) using the F-KPP model based on the knowledge of the visible tumor at day 90. In the middle image, a simple step-edge initialization while in the right image our tumor tail estimation is used. We observe that the prediction starting with an initial condition with appropriate tails is much more accurate in the case of a synthetic tumor grown by the F-KPP model.

The ability of the proposed formulation has been demonstrated using synthetic tumors. We are currently working on validating the performance of the formulation in images with histo-pathological data . Besides that, automatic estimation of model parameters, like D and ρ, from images taken at different time instances is an open problem we are working on.

References

1. Swanson, K., Alvord, E., Murray, J.: Virtual brain tumours (gliomas) enhance the reality of medical imaging and highlight inadequacies of current therapy. British Journal of Cancer **86** (2002)
2. Tracqui, P., Cruywagen, G., Woodward, D., Bartoo, G., Murray, J., Alvord, E.: A mathematical model of glioma growth: the effect of chemotherapy on spatio-temporal growth. Cell Proliferation **28**(1) (1995)
3. Clatz, O., Sermesant, M., Bondiau, P., Delingette, H., Warfield, S., Malandain, G., Ayache, N.: Realistic simulation of the 3d growth of brain tumors in mr images coupling diffusion with biomechanical deformation. IEEE T.M.I. **24**(10) (2005)

4. Jbabdi, S., Mandonnet, E., Duffau, H., Capelle, L., Swanson, K., Pélégrini-Issac, M., Guillevin, R., Benali, H.: Simulation of anisotropic growth of low-grade gliomas using diffusion tensor imaging. Magnetic Reson. in Med. **54** (2005)
5. Giese, A., Kluwe, L., Laube, B., Meissner, H., Berens, M., Westphal, M.: Migration of human glioma cells on myelin. Neurosurgery **38**(4) (1996)
6. Cristini, V., Lowengrub, J., Nie, Q.: Nonlinear simulation of tumor growth. Journal of Math. Biol. **46** (2003)
7. Patel, A., Gawlinski, E., Lemieux, S., Gatenby, R.: A cellular automaton model of early tumor growth and invasion. Journal of Theo. Biol. **213** (2001)
8. Murray, J.: Mathematical Biology. Springer-Verlag (2002)
9. Fisher, R.: The wave of advance of advantegous genes. Ann. Eug. **7** (1937)
10. Aronson, D., Weinberger, H.: Multidimensional nonlinear diffusion arising in population genetics. Advances in Mathematics **30** (1978)
11. Osher, S.: A level set formulation for the solution of the dirichlet problem for hamilton-jacobi equations. SIAM J. Math. Anal. **24**(5) (1993)
12. Bryson, S., Levy, D.: Central schemes for multidimensional hamilton-jacobi equations. SIAM J. Sci. Comput. **25** (2003)
13. McCorquodale, P., Colella, P., Johansen, H.: A cartesian grid embedded boundary method for the heat equation in irregular domains. J. Comp. Phys. **173**(2) (2001)

Simultaneous Stereoscope Localization and Soft-Tissue Mapping for Minimal Invasive Surgery

Peter Mountney[1], Danail Stoyanov[1], Andrew Davison[1], and Guang-Zhong Yang[1,2]

[1] Royal Society/Wolfson Foundation Medical Image Computing Laboratory,
[2] Department of Surgical Oncology and Technology
Imperial College, London SW7 2BZ, UK
{peter.mountney, danail.stoyanov, andrew.davison,
g.z.yang}@imperial.ac.uk
http://vip.doc.ic.ac.uk

Abstract. Minimally Invasive Surgery (MIS) has recognized benefits of reduced patient trauma and recovery time. In practice, MIS procedures present a number of challenges due to the loss of 3D vision and the narrow field-of-view provided by the camera. The restricted vision can make navigation and localization within the human body a challenging task. This paper presents a robust technique for building a repeatable long term 3D map of the scene whilst recovering the camera movement based on Simultaneous Localization and Mapping (SLAM). A sequential vision only approach is adopted which provides 6 DOF camera movement that exploits the available textured surfaces and reduces reliance on strong planar structures required for range finders. The method has been validated with a simulated data set using real MIS textures, as well as *in vivo* MIS video sequences. The results indicate the strength of the proposed algorithm under the complex reflectance properties of the scene, and the potential for real-time application for integrating with the existing MIS hardware.

1 Introduction

In surgery, the increasing use of MIS is motivated by the benefit of improved therapeutic outcome combined with reduced patient trauma and hospitalization. The technique is increasingly being used to perform procedures that are otherwise prohibited by the confines of the operating environment. MIS also offers a unique opportunity for deploying sophisticated surgical tools that can greatly enhance the manual dexterities of the operating surgeon. Despite the benefit of MIS in terms of patient recovery and surgical outcome, the practical deployment of the technique is complicated by the complexity of instrument control and difficult hand-eye coordination. Due to the large magnification factors required for performing MIS tasks, the field-of-view of the laparoscope cameras is usually very limited. This results in restricted vision which can affect the visual-spatial orientation of the surgeon and the awareness of the peripheral sites.

In order to facilitate the global orientation of the target site, a number of spatial localization techniques have been developed. These include the use of pre-operative imaging combined with 2D/3D registration such that the underlying structure and morphology of the soft-tissue can be provided. To cater for tissue deformation,

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 347 – 354, 2006.

structure from light [1] or motion sensors such as mechanically or optically based accelerometers [2, 3] are used. With the increasing availabilities of stereo-laparoscope cameras, detailed 3D motion and structure recovery techniques based on stereo vision have also be proposed recently [4, 5]. The major advantage of the optical methods is that they do not require additional modifications to the existing MIS hardware, and thus are easily generalizable to different clinical settings. One of the limitations of the above techniques is that they only consider information captured in the current field-of-view. Global information that is implicitly captured by the moving laparoscope camera is typically discarded. An exception to this is [6], where a map is built containing global information. The camera estimation is based on structure from motion, which is susceptible to drift.

The purpose of this study is to investigate the use of SLAM for simultaneous stereoscope localization and soft tissue mapping. In essence, the SLAM problem is concerned with the estimation of moving sensor while building a reconstruction of what it observes. The advantage of the method is that it builds a long-term map of the features with minimal drift, allowing localization of the sensor after long periods of feature neglect [7, 8]. This is particularly useful for laparoscope with restricted vision in that a global map of the operating field-of-view can be integrated with moving stereo vision. In this study, a sequential vision only approach is adopted which provides 6 DOF camera movement that exploits the available textured surfaces and reduced reliance on strong planar structures required for range finders. More importantly, it provides the potential for real-time application for integrating with the existing MIS hardware.

2 Methods

2.1 Building a Statistical Map

For stereoscope localization and soft tissue mapping, our aim is to recover the trajectory of the stereoscope and build a map of the environment. In a Kalman filter framework, the overall state of the system x is represented as a vector. This vector is partitioned into the state $\hat{x}_v$ of the camera and the states $\hat{y}_i$ of the features which make up the map. Crucially, the state vector is accompanied by a single covariance matrix which can also be partitioned as follows:

$$\hat{x} = \begin{pmatrix} \hat{x}_v \\ \hat{y}_1 \\ \hat{y}_2 \\ \vdots \end{pmatrix}, \qquad P = \begin{bmatrix} P_{xx} & P_{xy_1} & P_{xy_2} & \cdots \\ P_{y_1x} & P_{y_1y_1} & P_{y_1y_2} & \cdots \\ P_{y_2x} & P_{y_2y_1} & P_{y_2y_2} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{1}$$

The role of the covariance matrix is to represent the uncertainty to first order in all the quantities in the state vector. Feature estimates $\hat{y}_i$ can be freely added to or deleted from the map as required causing x and P to grow or shrink dynamically. In this framework, x and P are updated in two steps: 1) the prediction step uses a motion model to calculate how the camera moves during surgery and how its position

uncertainty increases; 2) the measurement step describes how the map and camera position uncertainties can be reduced when new input from the stereoscope is processed. Maintaining a full feature covariance matrix P allows the camera to re-visit and recognize known areas after periods of neglect (this has been irrefutably proven in SLAM research).

With the proposed approach, camera calibration is required to estimate 3D positions from stereo images and feature locations in the image plane from 3D positions. Calibration is done assuming a pinhole camera model and using a closed form solution [9]. The centre of the camera rig is taken to be the left camera and the extrinsic parameters describe the translation and rotation of the right camera relative to the left camera. In MIS, the stereoscopic laparoscope is pre-calibrated before the surgical procedure and remains unchanged.

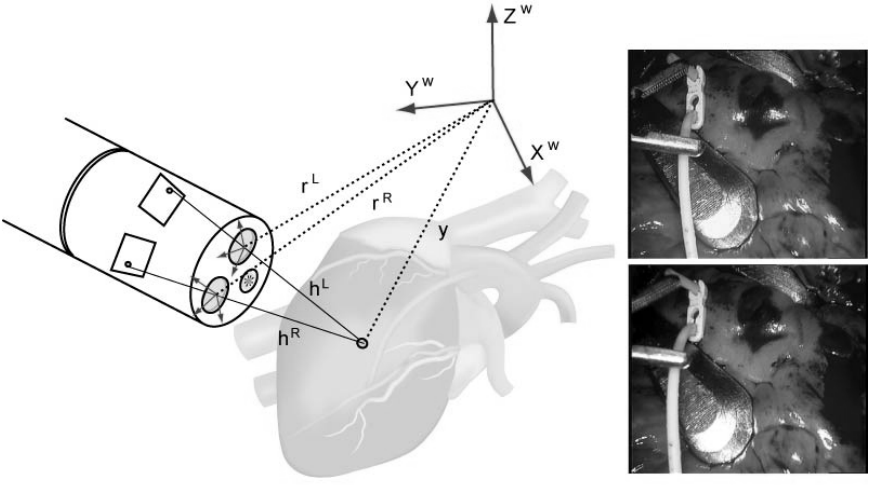


Fig. 1. Stereo-laparoscope camera geometry and an example image from a MIS scene. The figure illustrates the geometry between a global coordinate system, the local camera coordinates and a selected point from the map.

For the stereo-laparoscope camera, three coordinate frames illustrated in Fig. 1 are defined; W, fixed in the world, L, fixed with respect to the left camera and R, fixed with respect to the right camera.

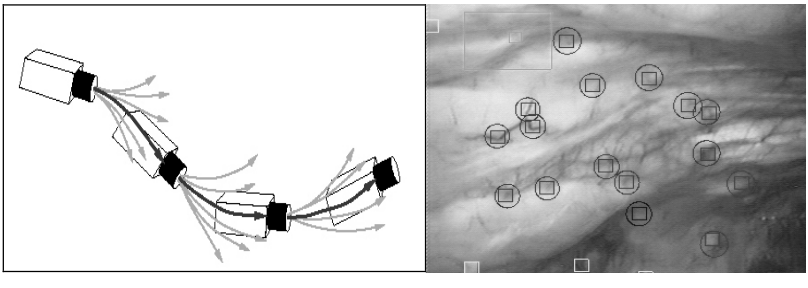
$$x_v = \begin{pmatrix} r^W \\ q^{WL} \\ v^W \\ \omega^W \end{pmatrix}, \qquad y_i = \begin{pmatrix} x_i^W \\ y_i^W \\ z_i^W \end{pmatrix} \tag{2}$$

We refer to x_v as the **state** of the camera rig. The state is made up of four parts; r^W a the position of the camera in the world coordinate system, q^{WL} the rotation of the camera in the world coordinate system, v^w is the linear velocity and ω^w the angular velocity. y_i refers to a feature consisting a 3D vector in XYZ Euclidean space.

2.2 Motion Model

In the case of a stereoscope moving during a MIS procedure, the motion model must take into account the unknown intentions of the operator. This unknown element can be modeled statistically by using a two part motion model. The first part is a deterministic element known as a "constant velocity, constant angular velocity model". This, however, does not mean that we assume that the camera moves at a constant velocity over all time. It only imposes that on average we expect its velocity and angular velocity to remain the same. The second part is stochastic and models the uncertainty in the surgeon's movement of the stereoscope. The uncertainty in the system is the acceleration modeled with a Gaussian profile. The implication of this model is that smoothness is implicitly imposed on the camera motion, very large accelerations are therefore relatively unlikely.

The rate of growth of uncertainty in this motion model is determined by the size of parameter P_n, and setting this to small or large values defines the smoothness of the motion we expect. With small P_n, we expect a very smooth motion with small accelerations, and would be well placed to track motions of this type but unable to cope with sudden rapid movements or changes in direction. High P_n means that the uncertainty in the system increases significantly at each time step, and while this gives the ability to cope with rapid accelerations the very large uncertainty means that a lot of good measurements must be made at each time step to constrain the estimates.



(a) **(b)**

Fig. 2. (a) Visualization of the model for 'smooth' motion: at each camera position a most likely path is predicted together with alternatives with small deviations. **(b)** A MIS scene where 25×25 pixel box represent feature patches detected using the Shi and Thomasi operator [10] and ellipses represent the estimated search regions for the landmarks.

2.3 Visual Feature

The *in vivo* anatomical structure is generally curved, thus making feature extraction more challenging than in man made environments. In [5], MSER features and weak gradient features were combined to create a dense 3D map of the heart. Features are tracked from frame to frame using a Lucas-Kanade tracker to recover the motion of the heart. These transient features work well for frame to frame tracking. However, in order to build a sustainable map, we require long term landmarks which are repeatable. A long term repeatable feature is one that is strongly salient and uniquely

identifiable. Previous work by Davison [6] has demonstrated long term features within a structured environment with a good degree of view point independence. This approach is used in this study to detect features using the saliency operator of Shi and Tomasi [9]. The feature is represented by a 25×25 pixel patch, and a normalized sum-of-squared difference correlation is used to match the feature in subsequent images. Specularities are removed through thresholding.

In the proposed framework, we aim to keep the number of visible features at a predetermined threshold to reduce reliance on weak features. A feature is "visible" if it is predicted to be in the current image. Features are added to the map if the number visible is less than this threshold. New features are detected in the left image and matched using normalized sum-of-squared difference in the right stereo image. Initialization is managed to prevent the same feature being tracked twice. Epipolar geometry is used to estimate the 3D position of the feature and all features are initialized with uniform uncertainty represented as a 3D Gaussian.

2.5 Measurement Model

Another important element of the proposed localization model is the measurement. The measurement model is the process for comparing the predicted SLAM map with the input from the stereoscope. The estimates x_p of camera position and y_i of feature position in 3D, allowing the position of the features to be predicted in the image plane. The position of a 3D feature relative to the camera is expected to be:

$$h_L^L = R^{LW}(y_i^W - r_L^W) \tag{3}$$

where R^{LW} is the rotation matrix transforming between the left camera frame L and world frame W. This is used to calculate (u_L, v_L) the predicted positions of the features in the left stereo image. The actual positions of the features in the images are obtained by actively searching the area around the predicted position. The search area is derived from the uncertainty of the feature's predicted position which is a 2D Gaussian p.d.f. over the image coordinates. Gating at three standard deviations provides an elliptic search window around the feature's predicted position.

3 Experimental Design

To validate the proposed method, a simulation with a virtual stereo camera moving through a texture mapped 3D world was rendered. The simulator as shown in Fig. 3 provides the ground truth data of known camera movement within a known environment. This allows the accuracy of the camera localization and mapping to be evaluated.

The camera motion was controlled so that the resultant inter-frame pixel motion did not exceed 20 pixels, which was consistent with observations from *in vivo* data. The virtual stereo camera rig was set up to replicate a stereo-laparoscope by taking similar camera intrinsic and extrinsic properties, notably the baseline was set to only 5mm. The environment contains a plane, which has been textured with an image taken from a robotic assisted totally endoscopic coronary artery bypass graft surgery

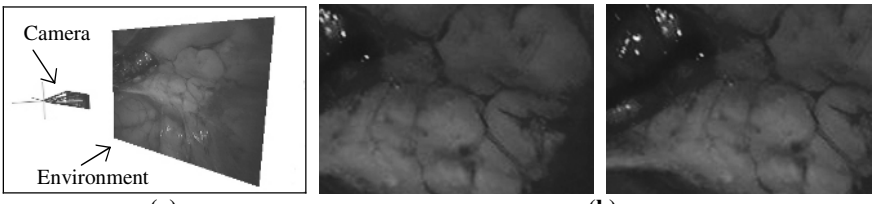


Fig. 3. An illustration of the simulation environment used to generate a stereo-laparoscopic video with known ground truth data for camera motion. A 3D rendition of the virtual world is shown in **(a)** and an example stereo pair taken from the virtual cameras is shown in **(b).**

to provide realistic image rendition. The use of a single planar model is not restrictive or degenerate as the proposed methods can be applied to more complex models.

In addition to synthetic simulations, the proposed technique has also been applied to real MIS videos. Since the ground truth data for the *in vivo* data is not available, qualitative analysis by forward tracking the motion and then reversing the video sequence is used to assess the internal consistency of the algorithm.

4 Results

In Fig. 4, the results of using the proposed technique to estimate the movement of the stereo-laparoscope over 176 frames of simulated video are provided. The stereo-laparoscope was moved by 1.5cm, 2cm and 0cm along the X, Y and Z axis respectively. Since no prior knowledge of the environment is taken, the initial estimations of feature positions have a large uncertainty. The uncertainty reduces as the stereoscope moves but creates a lag in the estimated movement. This is evident in the movement along the *X* axis. Small movement of around 1mm in the *Z* axis is a result of the narrow baseline of the stereoscope.

A potential problem with using a constant velocity motion models is the issue of dealing with sudden changes in direction. However, the results show the algorithm is robust to changes in direction. It can be shown that 87.7% of the recovered movement lies within three standard deviations of the ground truth, this represents a confidence interval of 99%.

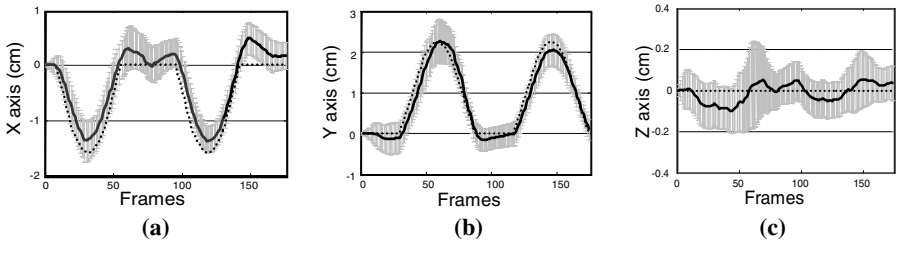


Fig. 4. Simulation based analysis of camera motion estimation. The graphs shown in **(a-c)** illustrate the recovered stereoscope movement in the *X*, *Y* and *Z* axes, respectively against the ground truth. The solid black line shows the estimated motion with the grey bars indicating the uncertainty associated with the estimate. The dotted line displays the ground truth motion.

For *in vivo* analysis, Fig. 5 shows example images from the left and right channels of the stereoscope, whereas Fig. 6 illustrates the recovered trajectory paths of the camera along the *X*, *Y* and *Z* axes in the world coordinate system. The original video footage is 79 frames and the reversed video is 79 frames long. It is evident from the graphs that the camera tracking closes the loop by returning the device close to its starting position. Finally, the SLAM map acquired from the *in vivo* sequence is shown in Fig. 7, along with an example image with selected features with their corresponding uncertainties. The appearance of features alters as the stereoscope and light source move. Feature matching is made more robust by using active search with the use of normalized sum-of-squared difference correlation to reduce data association errors.

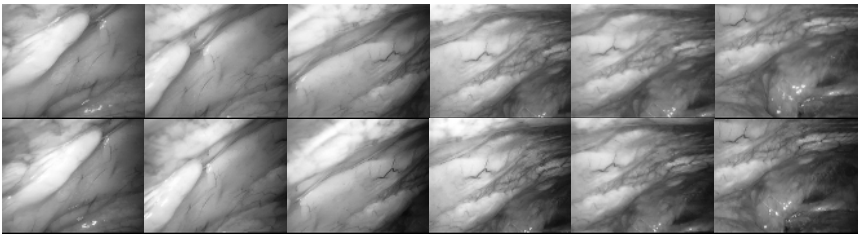


Fig. 5. Left (top) and right (bottom) stereo images taken from an *in vivo* stereo-laparoscope sequence that involves a change of camera viewing position and orientation

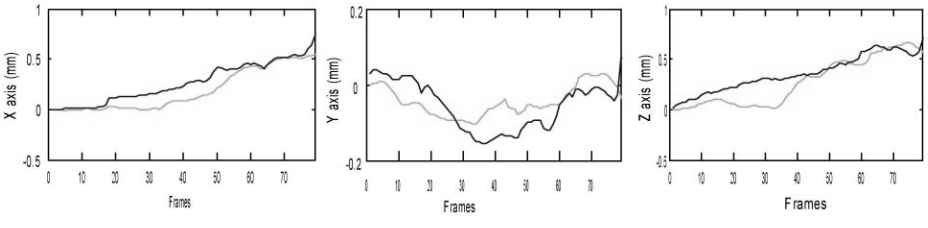


Fig. 6. *In vivo* analysis of the proposed techniques where the graphs show the recovered stereoscope movement along the *X*, *Y* and *Z* axes. Light grey lines represent the recovered motion in the forward sequence whereas the dark grey lines illustrate the recovered motion in the reverse direction.

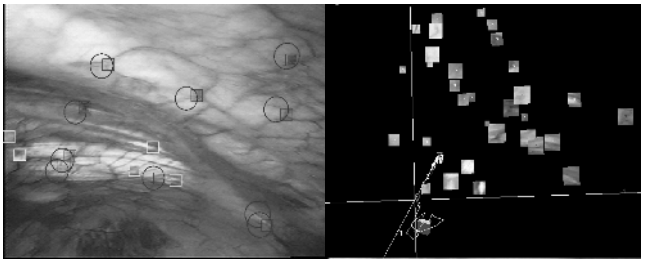


Fig. 7. Typical features selected in the left stereo image plane and the corresponding landmarks projected onto 3D coordinate system by using information built into the SLAM map

5 Discussion and Conclusions

In this paper, we have developed a technique to estimate the movement of the stereo-laparoscope during MIS and build a map of the anatomical structure. The method has been validated with a simulated data set using real MIS textures, as well as *in vivo* MIS video sequences. The results indicate the strength of the proposed algorithm under the complex reflectance properties of the scene. Accuracy can be further improved by incorporating information from the remaining stereo image into the measurement model, and directly cater for tissue deformation in the SLAM paradigm.

References

1. Keller K, Ackerman, J. Real-Time Structured Light Depth Extraction. in Proc of Three Dimensional Image Capture and Applications III SPIE, 2000, 11-18.
2. Hoffman J, Spranger M, Gohring D, Jungel M. Making Use of What You Don't See: Negative Information in Markov Localization. in Proc of Intelligent Robots and Systems, 2005.
3. Thrakal A WJ, Tomlin D, Seth N, Thakor N. . Surgical Motion Adaptive Robotic Technology (Smart): Taking the Motion out of Physiological Motion. in Proc of MICCAI, 2001, 317-325.
4. Chatila R, Laumond J. Position Referencing and Consistent World Modeling for Mobile Robots. in Proc of Robotics and Automation. 1985, 138-145.
5. Stoyanov D, Darzi, A., Yang, G-.Z. Dense 3d Depth Recovery for Soft Tissue Deformation During Robotically Assisted Laparoscopic Surgery. in Proc of MICCAI, 2004, 41-48.
6. Burschka D, Li M, Ishii M, Taylor RH, Hager GD. Scale-Invariant Registration of Monocular Endoscopic Images to Ct-Scans for Sinus Surgery. *Medical Image Analysis*, 2005, 9(5):413-439.
7. Davison AJ. Real-Time Simultaneous Localisation and Mapping with a Single Camera. in Proc of 9th IEEE ICCV, 2003, 1403.
8. Se S, Jasiobedzki, P. Instant Scene Modeler for Crime Scene Reconstruction. *IEEE A3DISS* 2005.
9. Zhang P, Milios EE, Gu J. Vision Data Registration for Robot Self-Localization in 3d. in Proc of Intelligent Robots and Systems, 2005, 2315-2320.
10. Shi J, Tomasi, C. Good Features to Track. in Proc of CVPR, 1994, 593-600.

Real-Time Endoscopic Mosaicking

Sharmishtaa Seshamani[1], William Lau[2], and Gregory Hager[1]

[1] Department of Computer Science, Johns Hopkins University, Baltimore, MD
{sharmi, hager}@cs.jhu.edu
[2] National Institutes of Health, Bethesda, MD*
william.lau@nih.gov

Abstract. With the advancement of minimally invasive techniques for surgical and diagnostic procedures, there is a growing need for the development of methods for improved visualization of internal body structures. Video mosaicking is one method for doing this. This approach provides a broader field of view of the scene by stitching together images in a video sequence. Of particular importance is the need for online processing to provide real-time feedback and visualization for image-guided surgery and diagnosis. We propose a method for online video mosaicking applied to endoscopic imagery, with examples in microscopic retinal imaging and catadioptric endometrial imaging.

1 Introduction

Endoscopy is an invaluable tool for surgical and diagnostic applications in pulmonary medicine, urology, orthopedic surgery and gynecology. It permits minimally invasive procedures, involving little or no injury to healthy organs and tissues. Current endoscopic technologies include fiberscopy, videoscopy, laparoscopy and wireless capsule endoscopy.

A drawback in these methods is the narrow field of view due to the size of most endoscopic imaging systems. As a result, individual images are often not very intuitive for evaluation. Automated mosaicking [1, 2] offers the opportunity to create an integrated picture or an environment map of a scene from a video sequence of endoscopic images. An essential first step in the process is the estimation of a registration estimate between captured images. One method of obtaining this estimate is the use of external optical tracking [3]. This however requires additional tracking equipment and a constant line of sight. A purely image based registration method is therefore an attractive alternative.

Current image registration methods generally apply to images related by planar homographies. Some examples are views of a plane from arbitrary camera positions and views of a general scene taken by a camera free only to pan, tilt and zoom [1]. Endoscopic images, however, are typically not related by planar homographies, due to the complexity of internal body scene structure and the impracticability of restricting camera motion. For example, bronchoscopy involves linear axial motion while imaging a tubular environment. It is therefore

* Disclaimer: The views and opinions of authors expressed herein do not necessarily state or reflect those of the NIH, DHHS, or the United States Government.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 355–363, 2006.

necessary to develop methods that, based on the imaging model and scene geometry, transform endoscopic images into a representation that is suitable for mosaicking. In particular, paracatadioptric imaging is potentially useful for endoscopy. A paracatadioptric system typically comprises of a parabolic mirror which reflects light onto a camera and thus provides a wider field of view which makes it a useful tool for endoscopy. However, image transformations imposed by the motion of a paracatadioptric imager are not linear, and further do not satisfy the requirement (for mosaicking) of forming a group, thus complicating the problem [4].

Methods for computing a registration estimate fall into two broad categories: direct methods which compute a transformation that optimizes some measure of photometric consistency over the entire image [5, 6], and feature based methods [7] which use a sparse set of corresponding image features to estimate the image-to-image mapping. Once a registration is computed, the construction of a mosaic entails resampling the images to a common coordinate system (given by the registration) so that they can be combined into a single image.

In most implemented systems, the entire mosaicking process is carried out offline, allowing the registration problem to be solved as a joint global optimization. However, this means that the quality of the mosaic and area it covers is difficult to determine until after the fact. A more intuitive approach is to develop an initial mosaic "online" as images are acquired. This provides the physician immediate and direct visual feedback as to the coverage and quality of the resulting mosaic.

In this paper, we describe methods for online image registration and mosaicking, and provide experimental results for retinal and endometrial imaging applications. In the next section, we first describe methods for performing traditional planar mosaicking and illustrate its application to retinal imaging. We then describe the modifications necessary to deal with paracatadioptric imaging of tubular structures and present results from that system.

2 Registration for Endoscopic Mosaicking

In the case of an endoscope viewing a planar or locally planar surface, the appropriate registration transformation is a homography [1]. However, in many cases the mapping between images of planar scenes can be described by mappings with fewer degrees of freedom which, consequently, can often be more reliably and rapidly estimated. In particular, affine mappings account for translation, rotation and scaling effects and are subgroups of a planar homography [1]. Although the affine transformation is, in general, necessary for mosaicking a large

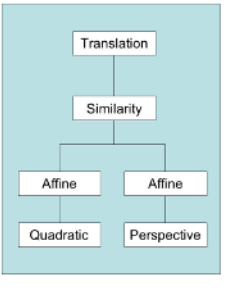


Fig. 1. Warping Model Hierarchy

scene, it is often possible to make due with even simpler models, allowing for simpler computation and a more stable result. Our mosaicking system begins with a simple translational model and moves through a hierarchy of models based on the scene structure (Fig. 1). The move to a more complex model is triggered when the registration error for the simpler model exceeds a fixed threshold. In the cases of locally planar (retinal) and cylindrical (endometrial) imaging which are presented in this paper, affine motion models generated small enough registration errors that did not exceed the threshold. Therefore quadratic and perspective models were not considered. The registration methodology we employ using this model is a direct technique. We denote the registration transformation as $D(p)$. In the case of an affine motion model, the transformation is linear which relates image coordinates as follows:

$$(u_1, u_2) = D(p)(x_1, x_2, 1)^T = f(x, p), \quad D(p) = \begin{pmatrix} 1+p_1 & p_3 & p_5 \\ p_2 & 1+p_4 & p_6 \end{pmatrix} \tag{1}$$

where $x = (x_1, x_2)$ is the pixel coordinate of a physical point on the first image, $u = (u_1, u_2)$ is the projection of the point in the second image, $p = (p_1, p_2, p_3, p_4, p_5, p_6)^T$ is the unknown parameter vector relating the two images and $f(x, p)$ is the transformation which is a function of x and p. [8, 5].

For online video mosaicking, motion between images is generally small, and the dominant motion is typically translation. As a result, it is common to compute an initial estimate of 2D translation by performing a brute-force search to maximize normalized cross-correlation between images. After this, a local continuous optimization method can be applied to compute the registration upto sub-pixel accuracy. Assuming brightness constancy, the goal of the registration algorithm is to minimize the sum of squared error between two images, I_0 and the image I_1 warped back onto the coordinate frame of I_0 ,with respect to the warping parameters p. For a general motion model with transformation function $f(x, p)$, this quantity is:

$$\sum_x [I(f(x, p + \Delta p)) - I_0(x)]^2 \tag{2}$$

This expression is linearized by a first order Taylor expansion on $I(f(x, p+\Delta p))$:

$$\sum_x [I(f(x, p)) + \nabla I \frac{\delta f}{\delta p} - I_0(x)]^2 \tag{3}$$

where ∇I is the image gradient vector and $\frac{\delta f}{\delta p}$ is the Jacobian of the transformation in (1). A linear closed-form solution can be obtained for the registration parameters p [6, 5]. In the interest of reducing computation time, a portion of the available image pixels can be chosen for optimization. In particular, since low magnitude image gradients have negligible effects on the solution, they can be eliminated to form an equivalent, smaller Jacobian matrix [8]. Once registration is computed to a subpixel level, the final step is to stitch warped images. The registration transformation between each pair of images j and k is defined as

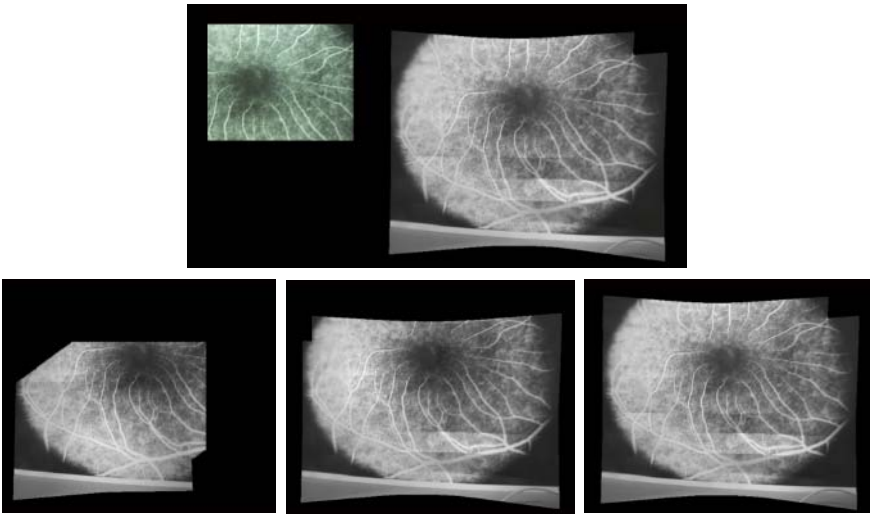


Fig. 2. An image mosaicking result for a retinal image sequence. The figure above shows an input image on the left and the resulting mosaic on the right; The 3 figures below show the mosaic in progressive stages. Note the change of field of view as more images are included.

$D(p)_{j,k}$. Define K_i as the transformation of an image to the mosaic. For each new image, the transformation $K_{i+1} = K_i.D(p)^{-1}_{i,i+1}$. The new image is then weighted and projected onto the current mosaic with this estimated transformation. The method has been tested on simulated retinal sequences, and on endoscopic retinal images acquired with a GRIN lens endoscope (Insight Imaging, Inc.). Fig. 2 shows an example of the former as it is being constructed, in real-time, using the methods described above. This mosaicker was implemented in C and runs at 30 frames/sec.

3 Warping Models for Catadioptric Imaging

Recently, we have begun to investigate mosaicking for paracatadioptric sensors moving in tubular structures. As noted previously, catadioptric images cannot be registered by transformations that form a group. However, for motion that is largely axial through a tubular structure which provides a scene of roughly constant depth (as is the case in several types of diagnostic endoscopy), paracatadioptric images can be transformed into cylindrical representations. The transformations between these cylindrical images satisfy the condition of group membership. Therefore these representations can be registered to form a "tubular" mosaic. Fig. 5 shows a phantom setup we have developed for illustrating such an imaging system. A catadoptric imager is mounted on a linear stage and is positioned inside an empty clear cylindrical tube. Different textures are affixed to the clear tube and are imaged by moving the camera steadily in a straight line

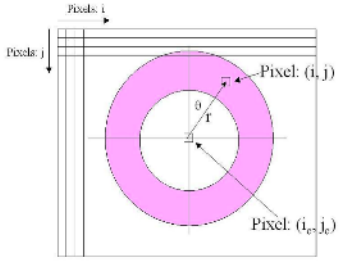


Fig. 3. Image Coordinates

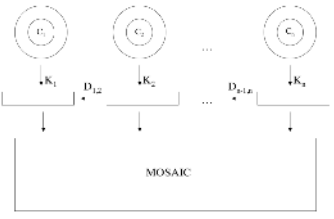


Fig. 4. Paracatadioptric Mosaicking

using the stage. This simulates the motion of an omnidirectional imager through different types of tissue.

A raw image from this sensor appears as an annulus with image coordinates (i_c, j_c) as its center (Fig. 3). Points (i, j) from the raw image can be described in polar coordinates (r, θ) by: $r_{i,j} = ((i - i_c)^2 - (j - j_c)^2))^{1/2}$ and $\theta_{i,j} = \tan^{-1}\left(\frac{i_c - i}{j_c - j}\right)$. For the imaging geometry used in our system, the polar representation can now be related to the cylindrical surface with coordinates $(x_{i,j}, y_{i,j})$ as follows:

$$x_{i,j} = r_{i,j}\theta_{i,j}, \quad y_{i,j} = \left(\frac{-(r_{i,j} - c)}{a}\right)^{1/b} \tag{4}$$

where a, b, and c are power law coefficients characterizing the imaging geometry.

Once the images are in cylindrical coordinates, axial motion becomes translation in y, and rotation about the imager axis becomes translational motion in x (now viewed as wrapping at the edges of the image). In order to apply the framework of the previous section, the paracatadioptric mosaicking algorithm requires the extra step of changing from radial to cylindrical coordinates (Fig. 4). Let $T(i, j) \mapsto (x, y)$ denote this change of coordinates and let J_T denote the Jacobian matrix of this transformation. It follows then that

$$\nabla_{x,y} I = J_T \nabla_{i,j} I \tag{5}$$

relates image gradients in the raw ((i, j) coordinates) image to those in the cylindrical ((x, y) coordinates) image. The latter can now be used in (3) to solve for the cylindrical motion between images using the raw image as it is acquired.

4 Results

A preliminary validation of the method was conducted using the large scale paracatadioptric camera system described in Section 3. The purpose of this validation was to measure the accuracy of the reconstructed mosaic.

The first required step was calibration, in order to determine power fit coefficients a, b and c from (6). To solve for these, a uniform grid affixed to

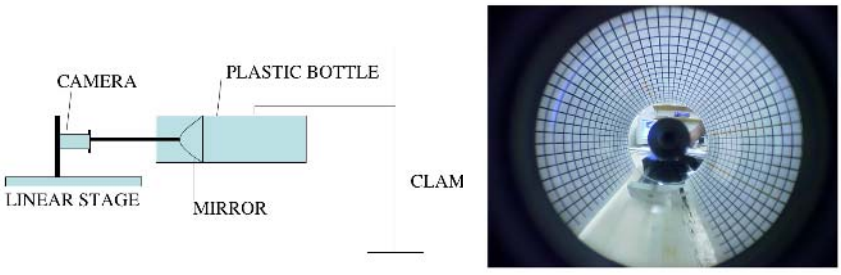


Fig. 5. Left:Large scale paracatadioptric simulator. Textures are affixed around the plastic bottle. The camera and mirror move in and out of the bottle during imaging. Everything else is stationary. Right:Calibration Grid.

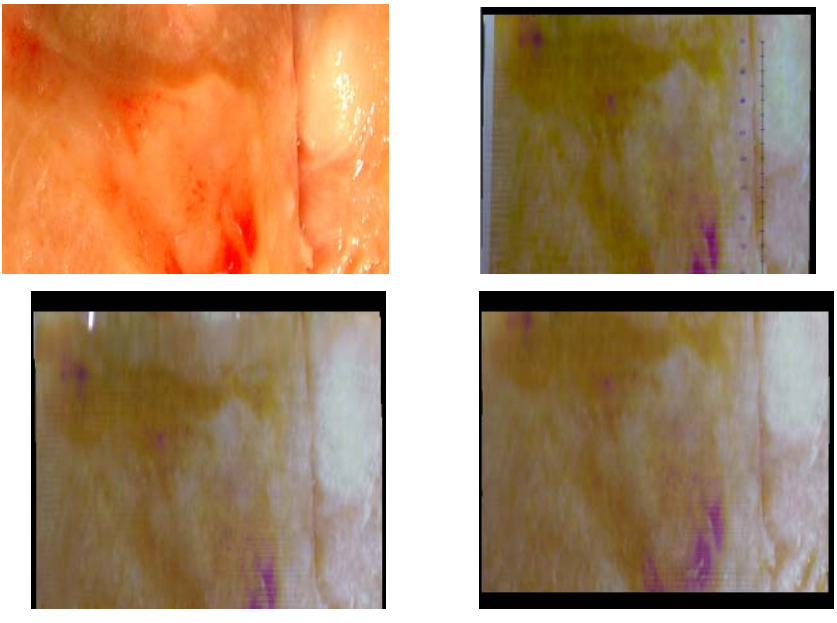


Fig. 6. Top Left: Original Myoma Texture Top Right: Mosaic of Myoma Texture with markers (ground truth) Bottom Left: Mosaic of Myoma Texture: Sample 1 Bottom Right: Mosaic of Myoma Texture: Sample 2

the imager was imaged and corners on the grid were automatically extracted (Fig. 5). Given this constant depth scene, the first grid circle could be chosen as the origin of the radial coordinate system allowing a solution for c. The distance between the first and second grid circle was taken as unit distance thus solving for a. Finally, taking the log of both sides yielded: $b\log(y) = \log\left(\frac{r-c}{a}\right)$. This linear relationship easily solved for b using additional grid points. Once b was computed, a and c could be recomputed by standard linear regression. The values

computed were as follows: a = 0.047, b = 1.51, c = 131.3. Tracking and mosaicking of uterine texture samples could then be implemented. Two types of data were collected. The first set of data was captured by imaging printed textures of normal and myomatous uterii affixed to the clear tube. For the second type of dataset, these printed textures were marked with pen mark fiducials placed at equal 1cm distances on a straight line. Image capture was then performed in the same manner as for the first type of dataset. The second set served as a validation tool. The uniformity of the reconstruction of the marks on the validation mosaics was used to determine the accuracy of the mosaicking algorithm. Fig 6(a) (top left) shows the original texture of a myomatous uterus which was pasted onto the tube. Fig 6(b) (top right) shows a reconstructed mosaic of the marked myoma texture. The distance between reconstructed marks ranged from 44 to 51 pixels (the image size is 551 X 661) and the variance of the distance between consecutive markers was 2.83 pixels. Given the ground truth distance of 1cm between markers, this gives a registration accuracy of 0.59mm. The ruler on the right side of the markers in Fig 6(b) provides a mark for every one half of the mean distance between reconstructed marks. Figs 6 (bottom two) show the mosaics generated from the dataset of the same texture without markers.

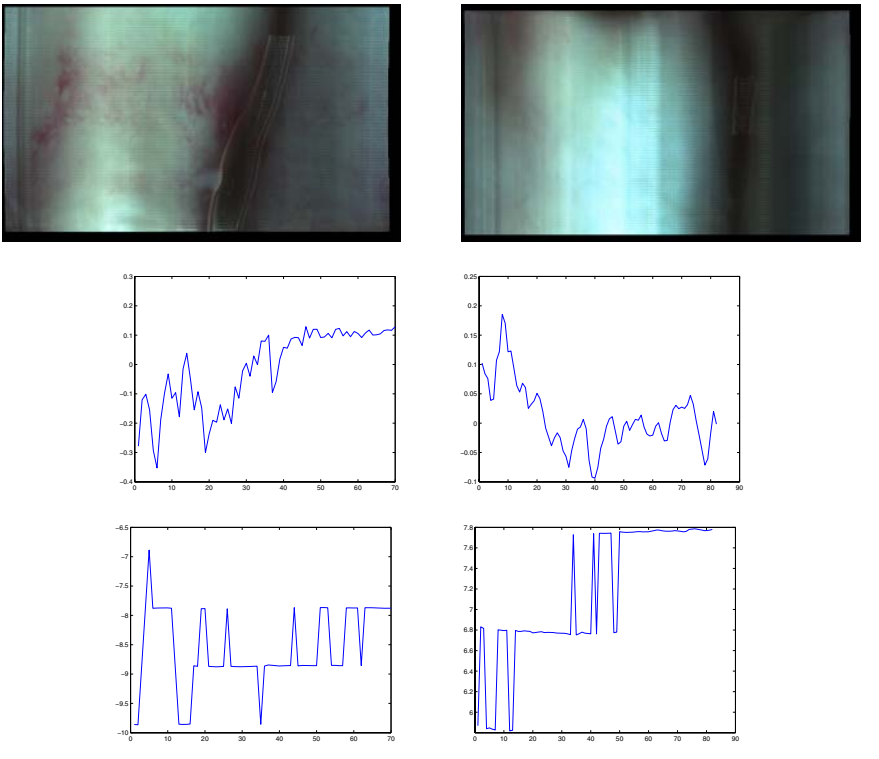


Fig. 7. Left and right columns show mosaics (first row), motion in x direction (second row) and motion in y direction (third row) of two different ex vivo samples

4.1 Experiments with Ex-vivo Data

The above method was then applied to ex-vivo uterus images. Data was captured using a 4mm diameter hysteroscope with a paracatadioptric imager. The hysteroscope was moved at uniform 1mm intervals between consecutive images. Sample 1 is a set of 70 images of an endometrium with a myoma. Sample 2 is a set of 83 images of a normal endometrium. The resulting mosaics and motion plots are shown in Fig. 7. The computed motion in the y direction which corresponds to the dominant translational motion varies between 7 and 10 pixels between image pairs in the two samples. This is in accordance with the constant, purely translational motion of the imager. The computed motion in the x direction is very small as expected.

5 Discussion

We have presented a novel approach for online image tracking and mosaicking for the improved visualization of locally linear surfaces and closed tubular environments in the cases of microscopic and omnidirectional contact imaging. The ability to perform real-time processing is particularly important in the case of endoscopic mosaicking in order to provide real-time visualization. In general, image mosaicking is subject to some level of drift as there is no way to completely eliminate small incremental motion estimation errors. In order to reduce these registration errors a global block adjustment alignment can be applied to the whole sequence of images in offline processing, resulting in an optimally registered mosaic [2].

Future work will focus on extending the method to interpret lateral motion of the catadioptric imager, incorporation of tissue deformation and further ex-vivo experiments.

Acknowledgments

This research was supported by the National Institutes of Health and the CISST Engineering Research Center under the National Science Foundation Grant No. EEC9731748.

References

1. Capel, D.: Image Mosaicing and Super-resolution. PhD thesis, Robotics Research Group, Department of Engineering Science, University of Oxford (2001)
2. Shum, H., Szeliski, R.: Construction of panoramic image mosaics with global and local alignment. IJCV **16** (2000) 63–84
3. D.Dey, Gobbi, D., an K.J.M Surry, P.S., Peters, T.: Automatic fusion of freehand endoscopic brain images to three-dimensional surfaces: Creating stereoscopic panoramas. IEEE Transactions on Medical Imaging **21** (2002) 23–30

4. Dornaika, F., Elder, J.H.: Image registration for foveated omnidirectional sensing. In: ECCV (4). (2002) 606–620
5. Baker, S., Matthews, I.: Lucas-kanade 20 years on: A unifying framework: Part 1. Technical report, CMU-RI (2002)
6. Hager, G., Belhumeur, P.: Efficient region tracking with parametric models of geometry and illumination. IEEE PAMI **20** (1998) 1025–1039
7. Can, A., Stewart, C., Roysam, B., Tanenbaum, H.: A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. IEEE PAMI **24** (2002) 347–364
8. Lu, L., Dai, X., Hager, G.: Real time video mosaicing. cirl technical report. Technical report, Dept of Computer Science, Johns Hopkins University (2003)

Depth Perception – A Major Issue in Medical AR: Evaluation Study by Twenty Surgeons

Tobias Sielhorst[1], Christoph Bichlmeier[1], and Sandro Michael Heining[2], and Nassir Navab[1]

[1] Chair for Computer Aided Medical Procedures (CAMP), TU Munich, Germany
{sielhors, bichlmei, navab}@cs.tum.edu
[2] Chirurgische Klinik und Poliklinik - Innenstadt, LMU Munich, Germany
sandro-michael.heining@med.uni-muenchen.de

Abstract. The idea of in-situ visualization for surgical procedures has been widely discussed in the community [1, 2, 3, 4]. While the tracking technology offers nowadays a sufficient accuracy and visualization devices have been developed that fit seamlessly into the operational workflow [1, 3], one crucial problem remains, which has been discussed already in the first paper on medical augmented reality [4]. Even though the data is presented at the correct place, the physician often perceives the spatial position of the visualization to be closer or further because of virtual/real overlay.

This paper describes and evaluates novel visualization techniques that are designed to overcome misleading depth perception of trivially superimposed virtual images on the real view. We have invited 20 surgeons to evaluate seven different visualization techniques using a head mounted display (HMD). The evaluation has been divided into two parts. In the first part, the depth perception of each kind of visualization is evaluated quantitatively. In the second part, the visualizations are evaluated qualitatively in regard to user friendliness and intuitiveness. This evaluation with a relevant number of surgeons using a state-of-the-art system is meant to guide future research and development on medical augmented reality.

1 Introduction

Real-time in-situ visualization of virtual data overlaid on its real counterpart is getting increasing attention. Augmented Reality (AR) for intra-operative visualization and navigation has been a subject of intensive research and development during the last decade. Birkfellner et al. have realized an optical see-through, [3] in-situ visualization system based on an operating binocular that is used for maxillofacial surgery. Sauer et al. [5] realized in-situ visualization with a video see-through technology head mounted display (HMD). Edwards et al. have developed an operating microscope with in-situ visualization capabilities for brain surgery [6]. These research groups report in their late works that their systems meet the desired accuracy in terms of tracking and overlay. Edwards and his colleagues report unstable and ambiguous depth perception despite accurate system calibration [7]. They report misperceptions due to multiple transparent

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 364–372, 2006.

surfaces with textures. Already in the first paper on medical AR, where Fuchs et al. described a video see-through system for overlaying echography images, the authors discussed the issue of depth perception in the scene. They have superimposed the virtual images on the real image. Unfortunately, this visualization appears to be in front of the patient, because the image occludes the patient. Therefore, the depth cue of occlusion tells the brain that it is in front of the patient even though the images are correctly rendered in terms of position. Psychologists distinguish in the field of cognition and perception [8] sixteen different depth cues that are exploited by the human visual system for depth perception. These depth cues have different persuasive power, they have different precision and they interact in different ways with each other. Furthermore, some depth cues have absolute quality while others only have relative quality. This means that visual system can estimate from the accommodation of the eye an absolute value for the distance of an object. However, the depth cue of occlusion can only tell, which one of two objects is in front. Since the depth cues have such a complex behavior, a straightforward construction of a perfect visualization is not possible. Apart from the correct depth perception, a compelling visualization should allow for easy interaction. That means that surgeons can understand the visualized data quickly and change the visualization easily according to the desired information. Our evaluation intends to bring more knowledge into the community about the influence of visualization techniques on depth perception. Among the tested visualizations, we have integrated the ones mentioned above from other research groups, and we also suggest new kinds of visualizations.

2 Methods

We have compared seven different visualizationmodes relative to each other with the same hardware setup. In the following paragraphs, we describe our hardware setup, the evaluation procedure and the visualization modes that we have evaluated

2.1 Description Technology

Figure 1 gives a complete overview about an AR system in use. The hardware setup is similar to the one proposed by Sauer et al. [2] for medical augmented reality. The augmentation is provided by a stereoscopic video see-through HMD (figure 3, bottom) worn by the surgeon. The virtual images are merged with two color cameras; one for each eye. Video see-through technology offers for our experiments a broader range of visualization possibilities than optical see-through, since virtual objects may be displayed opaque.

For superior registration quality, the system uses two synchronized tracking systems. The single camera inside-out tracking system allows for a high rotational precision, which is necessary for tracking of the HMD. The optical outside-in tracking system with four cameras fixed to the ceiling covers a large working area. Both of the systems use the same kind of retroreflective fiducial markers offering a registration free transformation from one tracking system to

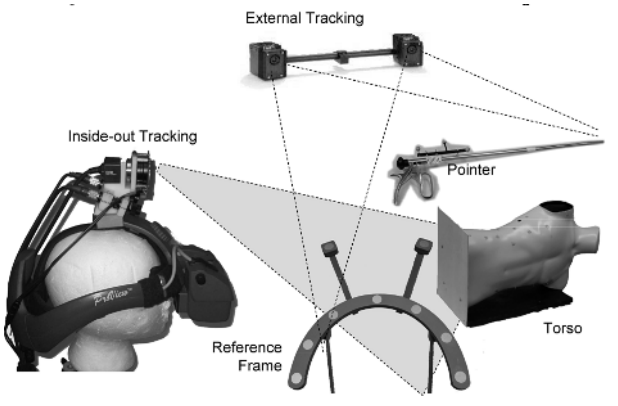


Fig. 1. Technical overview of the experimental setup

the other. These fiducial markers are attached to surgical instruments as well as on the body phantom (see figure 1). In order to recover the six degrees of freedom of a rigid body, the external optical tracking system requires four rigidly attached markers. As a third tracked object in the scene, there is a common reference target that facilitates the fusion of the inside-out and the outside-in tracking systems with no need for hand-eye coordination. The off-the-shelf computer used to render 3D graphics, to compute and include tracking data, to synchronize and combine imagery of virtual and real entities is an Intel Xeon(TM), CPU 3,20 GHz, 1,80 GB RAM, NVIDIA Quadro FX 3400/4400.

2.2 Description of the Evaluation Procedure

The evaluation of the visualization methods is divided into two parts. In the first part, we want to evaluate, which visualization offers the most reliable depth perception. In order to get quantitative results, we let the participants fulfill a task on a phantom and measure the accuracy of the performance and the time to fulfill the task.

In the second part, we survey the user acceptance. The participants fill out a questionnaire regarding the usability of the system. This questionnaire has been filled out directly after the first part of the evaluation. Thus we expect the experience with the AR system to be very present and pristine. The group of participants consists of 20 surgeons of our partner hospital.

2.3 Description of the Tests

All participants had to fulfill the same task with different visualization techniques. The task and the visualization modes can be viewed in the video[1]. The task consists of a pointer that has to be moved to a spot inside a body phantom.

[1] http://campar.in.tum.de/files/publications/Sielhorst2006DepthPerception.video.avi

The body phantom consists of a plastic surface of a torso and a spinal cord model inside. The data model has been recorded with a CT scanner before.

During the test, each participant has to move a pointer to randomly located spots on the surface of the spine of the phantom as shown in figure 2. Each participant had to find twenty points for each visualization. For the tests, the model of the spine has been taken out of the body phantom in order not to get haptic feedback. The participants have to rely on their mere vision. The phantom guarantees that the participants do not have a direct view to the point of interest, which is also true for minimally invasive surgery. The participants have been asked to move the pointer to the target spot and tell if they are done. The spot has been marked with a small black point in order not to interfere with the current visualization (figure 2).

For each participant we provided two minutes during he or she got used to the HMD and AR visualization. After this, the participants were asked to fulfill the tasks above. The average time for the whole test was 16.4 minutes. This exceeds the expected usage time of the visualization system in the operating room.

Fig. 2. Augmented surgical instrument serves as the pointer device. The small black point is visualized. The surgeons were asked to touch the point with its tip in the virtual position.

2.4 Evaluated Visualization Modes

The following visualization modes have been tested in a random order in order to average out effects of learning and exhaustion. The table in figure 4 contains the visualizations in the same order as figure 3. Explanation of these follow in the order of figure 3. The visualization speed is denoted in brackets in frames per second.

Triangle mesh. The triangle mesh (see figure 3.4) is a representation of the surface of the bone structure in the scene. The surface has been segmented from the CT scan before the evaluation. The surface is stored in the computer as a list of triangles. In this mode only the edges of the triangles are displayed. Edges

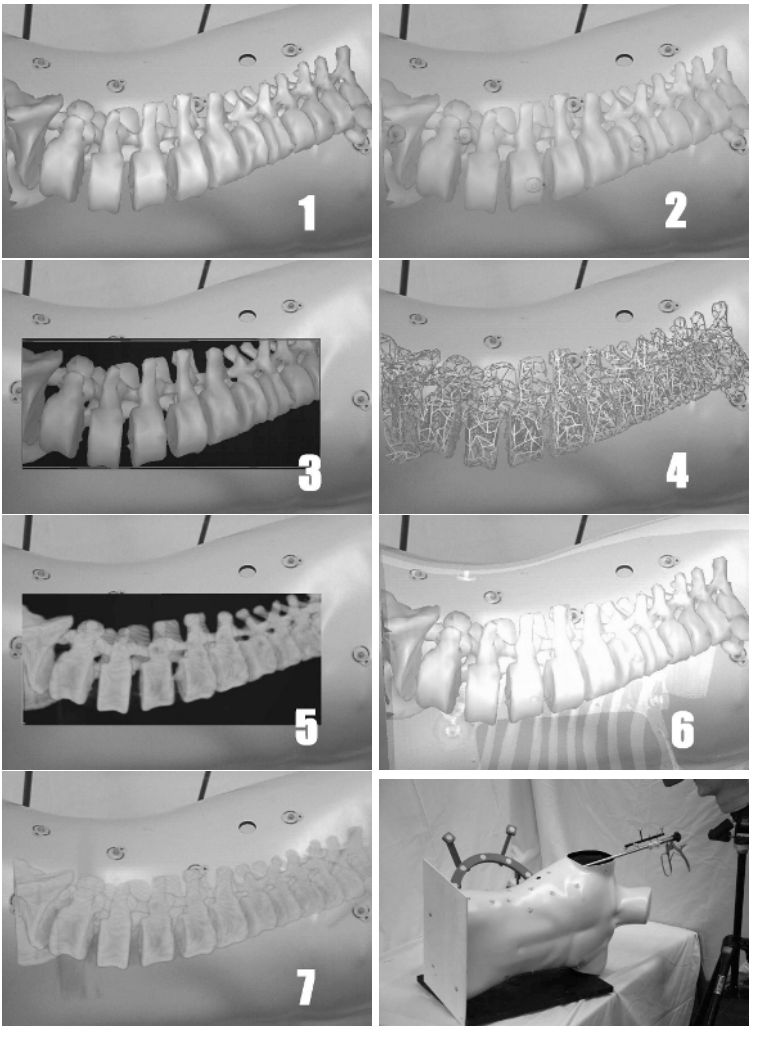


Fig. 3. The seven evaluated visualizations; bottom right: The setup

in a higher distance to the viewer are displayed darker than closer ones in order to give a strong depth impression.

Surface rendering. The surface rendering (see figure 3.1,2,3,6) is as the triangle mesh a representation of the surface of the bone structure in the scene. The surface has been also segmented from the CT scan. The surface is visualized with untextured, but shaded solid triangles.

Volume rendering. Volume rendering (see figure 3.5,7) represents the whole volume rather than a surface. The data need not be segmented or prepared. Each voxel is rendered with a certain transparency according to its value. The transfer function that relates the value to its color and transparency has been chosen in

	Depth perception	Effective-ness	Qustionaire: Experience of Perception (1 = good, 5 = bad)	Frame-rate in our system
Surface rendering opaquely superimposed	o	+	2,2381	30 fps
Surface rendering transparently superimposed	+	+	2,5238	30 fps
Surface rendering through a virtual window in the skin	+	+	2,4762	30 fps
Triangle mesh	-	+	3,4211	30 fps
Volume rendering model through a virtual window in the skin	-	-	2,2857	9-10 fps
Surface rendering with a glass effect of the skin	-	-	2,8571	7-9 fps
Volume rendering superimposed	-	-	2,381	5-6 fps

Fig. 4. Table of tested visualizations including simplified results

a way that bone structure in the spine model is emphasized. The rendering is performed with the support of 3D texture hardware as suggested by Hastreiter et al. [9]

Glass effect. This effect is only applied to the surface of the skin, but not the vertebrae (see figure 3.6). The skin is rendered transparently and achromatically. Only reflections of a virtual light source to the skin are rendered. By this means the skin looks like glass. The idea is to provide a visual impression that the skin is in front of the bone model, since the reflections of the glass skin occlude slightly the bone model or influence its colour at certain areas.

Virtual window. This effect is also only applied to the skin(see figure 3.3,5). In this visualization add-on we define a region on the skin, which we call the virtual window. The visualization of the bone can only be seen through the virtual window, but not outside. The window enables the effective depth cues *occlusion* and *motion parallax* because its frame occludes partially the spinal column in the background and the projections on the retina of these stationary objects move relative, which is caused by observer movement.

3 Results

We have evaluated an overall number of 2398 3D points that had to be touched in the model. We would like to point out that this evaluation is not intended to find out the accuracy of the system. Especially the exaggerated length of the pointer (64 cm) and its crude tip with a diameter of 9 mm has been designed to provoke errors due to interaction. A shorter pointer that is peaked at the end yields far more accurate results, but it would emphasize errors of the system and unwanted movements like tremor, which we would like to totally disregard.

The volume rendering visualization proved to have a performance that can be displayed in a current PC based AR system even with a large data volume. However other kinds of volume visualizations that need prior segmentation showed

a faster and more precise interaction. Even though the surgeon liked the volume rendering and reported a good perception of depth (see figure 4).

The triangle mesh mode did not perform as well as expected. The reason might be that the visualized model has a complex structure of many self occluding surfaces that confuse the viewer when rendered with transparent triangles.

Figure 5 (top right) shows the performance in the course of time. The participants were able to work with the video see-through HMD without losing performance during the whole experiment that took 16.4 minutes in average. The graphs show no indication of exhaustion during the test. The accuracy and speed of interaction improved over the whole experiment. In the beginning of the experiment the participants trade accuracy for interaction speed (see figure 5, top right). We interpret this as indicator for getting used to the test and to the unfamiliar system.

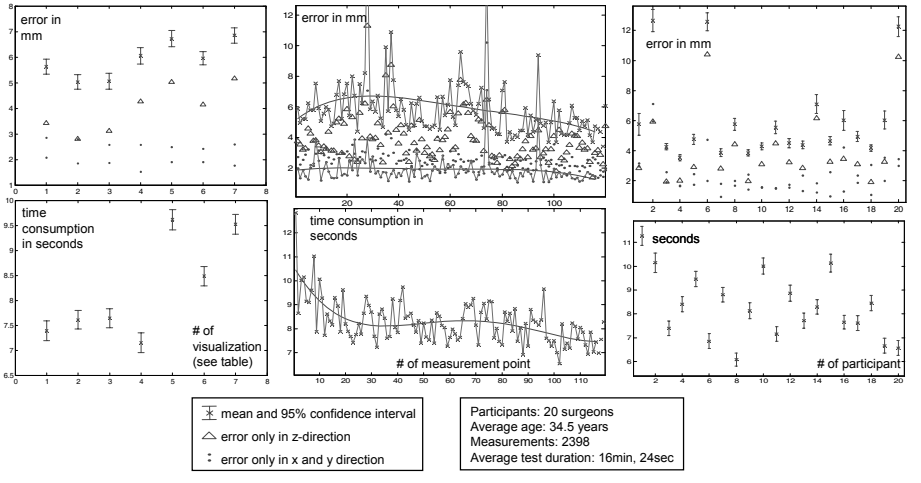


Fig. 5. Figures of the evaluation. Visualization 2 and 3 conveys depth best(left), fast visualizations cause fast user performance; users try to compensate missing depth with lateral accuracy, user trade speed for accuracy during the first 30 points i.e. relaxing after new situation (middle); even after 110 points (-$\frac{1}{4}$ hour) participants get still faster while becoming more accurate i.e. concentration does not drop; deviations of the partipicants: Fast interaction does not necessarily mean bad results and vice versa (right)

Figure 5 (top left) shows the performance of each visualization mode. The numbers are according to figure 3 and 4. The different visualization modes have a clear impact on the interaction precision and speed. In our experiments the frame rate seems to be an important factor for the performance. The participants complained about slow visualization modes and had slower and less precise performance using these. The graph shows the average error as well as the error splitted up into x and y and the z-axis that is parallel to the viewing direction. The fast modes using surface rendering clearly show better results than the fast modes using triangle mesh. It also performs better than the slow modes using volume rendering or surface rendering. Therefore, we suggest using the two best

visualization modes for interacting with a 3D model in medical AR. This is the transparent surface rendering or the one with the virtual window. It would be quite exiting to find out, if a combination of these could further improve the performance.

Figure 5 (bottom left) shows the performance of each individual surgeon. Comparing the graph depicting speed and the accuracy, we can see that these do not correlate with each other. That means that the success depends on the skills and not on the speed of the performance.

4 Conclusion

We conclude that the kind of visualization makes a serious difference in terms of effectiveness with the interaction. These findings do not even include the fact whether a visualization offers the desired medical information. This is application dependent. We have only been interested in the depth perception with different visualizations for stereoscopic augmented reality in the medical context. Due to hardware developments in the recent five years we can offer new visualization techniques that have been compared to the classic ones. One of the new visualization could outperform the other tested visualizations in terms of depth perception and effectiveness. This supports furthermore our thesis that visualization and interaction with augmented reality have to be better understood for medical application to take dvantage of its power.

Acknowledgments. We would like to thank each of the participants of our clinical partner for their time and their comments in the evaluation. Without their commitment the whole evaluation would not have been possible. We thank Siemens Corporate Research Princeton, NJ, USA for the courtesy of their AR system RAMP. In particular we thank Frank Sauer and Ali Khamene for continuous communication.

References

1. King, A.P., Edwards, P.J., Maurer, et al.: A system for microscope-assisted guided interventions. IEEE Trans. Med. Imag. **19** (2000) 1082–1093
2. Sauer, F., Khamene, A., Vogt, S.: An augmented reality navigation system with a single-camera tracker: System design and needle biopsy phantom trial. In: MICCAI 2002. Volume 2489. 116–124
3. Birkfellner, W., Figl, M., Huber, K., et al.: A head-mounted operating binocular for augmented reality visualization in medicinedesign and initial evaluation. IEEE Trans. Med. Imag. **21** (2002) 991–997
4. Bajura, M., Fuchs, H., Ohbuchi, R.: Merging virtual objects with the real world: seeing ultrasound imagery within the patient. In: Proceedings of the 19th annual conference on Computer graphics and interactive techniques, ACM Press (1992) 203–210

5. Sauer, F., Khamene, A., Bascle, B., Vogt, S., Rubinob, G.J.: Augmented reality visualization in imri operating room: System description and pre-clinical testing. In: Proceedings of SPIE, Medical Imaging. Volume 4681. (2002) 446–454
6. King, A., Edwards, P., Maurer, et al.: Stereo augmented reality in the surgical microscope. Presence: Teleoperators and Virtual Env. **9** (2000) 360–368
7. Johnson, L.G., Edwards, P., Hawkes, D.: Surface transparency makes stereo overlays unpredictable: The implications for augmented reality. In: Medicine Meets Virtual Reality (MMVR). IOS Press (2002) 131–136
8. Cutting, J.E., Vishton, P.M.: Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In: W. Epstein & S. Rogers (Eds.), Perception of Space and Motion. (1995) 69–117
9. Hastreiter, P., Rezk-Salama, C., Tomandl, B., Eberhardt, K., Ertl, T.: Fast analysis of intracranial aneurysms based on interactive direct volume rendering and cta. In: MICCAI '98: 660–669

Hybrid Navigation Interface for Orthopedic and Trauma Surgery

Joerg Traub[1], Philipp Stefan[1], Sandro Michael Heining[2], Tobias Sielhorst[1], Christian Riquarts[2], Ekkehard Euler[2], and Nassir Navab[1]

[1] Chair for Computer Aided Medical Procedures (CAMP), TU Munich, Germany
{traub, stefanp, sielhors, navab}@cs.tum.edu
[2] Chirurgische Klinik und Poliklinik - Innenstadt, LMU Munich, Germany
{Sandro-Michael.Heining, Christian.Riquarts, Ekkehard.Euler}@med.uni-muenchen.de

Abstract. Several visualization methods for intraoperative navigation systems were proposed in the past. In standard slice based navigation, three dimensional imaging data is visualized on a two dimensional user interface in the surgery room. Another technology is the in-situ visualization i.e. the superimposition of imaging data directly into the view of the surgeon, spatially registered with the patient. Thus, the three dimensional information is represented on a three dimensional interface. We created a hybrid navigation interface combining an augmented reality visualization system, which is based on a stereoscopic head mounted display, with a standard two dimensional navigation interface. Using an experimental setup, trauma surgeons performed a drilling task using the standard slice based navigation system, different visualization modes of an augmented reality system, and the combination of both. The integration of a standard slice based navigation interface into an augmented reality visualization overcomes the shortcomings of both systems.

1 Introduction

Standard slice based navigation systems are commonly used and commercially available for orthopedic and trauma surgery. In general they consist of a position and orientation tracking system and a two dimensional user interface. These systems visualize the navigation information based on three dimensional medical imaging data on an external monitor. The three major drawbacks of state of the art navigation systems are a) every imaging and navigation device comes with its own user interface, b) the guidance information based on three dimensional data is visualized on two dimensional user interfaces, and c) the navigational information is not visualized directly on the operation situs, forcing the surgeon to observe the navigation information at a different location as the action is performed.

Augmented reality visualization was introduced as an alternative user interface for navigated surgery. The navigation information is superimposed onto the surgeon's view of the real world. In the past decade numerous applications and hardware setups using augmented reality visualization in medical navigation

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 373–380, 2006.

were proposed. King, Edwards et al. [1] developed a stereoscopic system called MAGI that is used for microscope based neurosurgery. Birkfellner et al. [2] designed and developed the VarioscopeAR, an augmented reality head mounted operation microscope, for maxillofacial surgery. Sauer et al. [3] use a stereoscopic video-see through head mounted display to visualize three dimensional imaging data for various domains in interventional radiology and surgery [4].

In the past standard navigation systems and augmented reality were presented as concurrent approaches and evaluations dealt with the comparison of two dimensional user interfaces versus three dimensional in-situ visualization techniques [5]. Based on a head mounted display augmented reality system (section 2) we implemented a hybrid navigation interface (section 3.3) as a combination of the standard slice based navigation (section 3.1) and augmented reality visualizations (section 3.2). We propose to fuse the two separate interfaces to create one single three dimensional user interface for orthopedic and trauma surgery applications. Exemplary applications are pedicle screw placement in spinal surgery, orthopedic implant positioning and osteochondritis dissecans. In order to evaluate possible benefits of the fusion of these systems isolated systems and its usefulness for surgery, we designed a phantom experiment, in which a drill must be navigated to a given target location. Three surgeons evaluated different visualization systems through a set of detailed experiments (section 4).

2 System Setup

Our system is a combination of two existing components, a three dimensional user interface and an optical tracking system. We developed a fully automatic procedure that is based on a CT scan with attached fiducials that are visible in CT and tracking space. This allows to use the proposed visualization on any object after a CT scan with attached fiducials is performed.

2.1 Hardware Setup

The augmented reality system is based on a stereoscopic video see-through head mounted display (HMD) developed by Sauer et al. (SCR, Princeton, USA) [3]. The head mounted display is equipped with two color cameras to obtain images of the observed scene. Additionally a tracking camera is attached to the system for head pose estimation[6]. This technique, often referred to as inside-out tracking, has been proven to minimize the error of visualization in augmented reality [7].

There are two reasons for the preference of a video-see-through display to an optical-see-through device. Firstly these systems achieve a perfect synchronization of video and head pose data since the cameras are genlocked, eliminating any time lag between the images of the cameras, which could lead to perceivable jitter or swimming [8]. Secondly we have more options for visualization since we have the full control over the image display, whereas in optical systems only brightening of the augmentation is possible.

The drawback of using a single camera for instrument tracking is that large marker configurations are needed for a precise localization of targets[9]. As large

tracking targets attached to instruments are not desired in the operating theatre, we use the ARTtrack1[1] external optical tracking system to track instruments and surgical objects. This device is capable of tracking targets within our working volume with an accuracy of < 0.2 $[mm]$ RMS. A marker frame (Fig. 1(D)) is used to establish a common coordinate frame between the inside-out and external optical tracking system. The entire hardware setup is shown in Fig. 1.

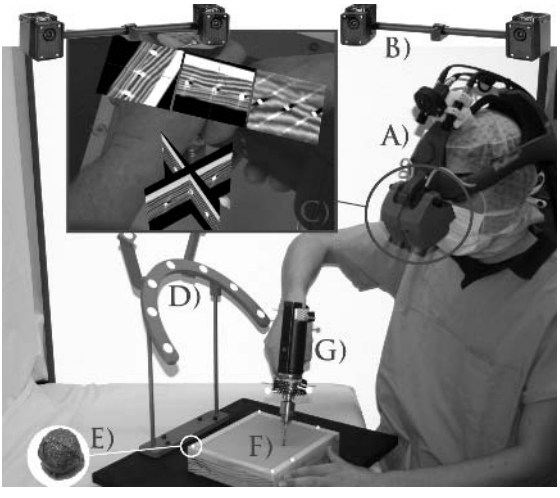


Fig. 1. Illustration of the setup for the hybrid navigation interface. (A) The HMD with two cameras for the video images and a single camera tracker for determination of the pose relative to the marker frame (D). An external optical infrared tracking device (B) is used for tracking surgical instruments (G) and CT detectable, infrared retro-reflective markers (E) attached to the phantom (F). The hybrid navigation view (F) is displayed on two miniature LCD monitors. In this augmentation all coordinate systems are visualized representing the transformations involved. The marker frame (D) is used as a common coordinate system for both, single camera tracking (A) and external optical tracking (B).

The transformation from the coordinate system of the external tracking device to two dimensional coordinates in the overlay image is given by

$$^{Overlay}H_{Target} = {}^{Overlay}H_{Cam}\ {}^{Cam}H_{Frame}\ \left({}^{Ext}H_{Frame}\right)^{-1}\ {}^{Ext}H_{Target} \quad (1)$$

where the transformations ${}^{Ext}H_{Frame}$ and ${}^{Ext}H_{Target}$ are provided by the external tracking system, ${}^{Cam}H_{Frame}$ and ${}^{Overlay}H_{Cam}$ are derived using Tsai calibration.

2.2 Clinical Integration

The requirement for a navigation system applicable in trauma surgery is a seamless integration into the clinical workflow and an automatic configuration with no additional interactive calibration or registration procedures during surgery.

[1] A.R.T. GmbH, Weilheim, Germany.

Most methods described in literature use tracked pointers to register markers in patient space with their corresponding centroids segmented from imaging data [1]. We designed markers that are automatically detectable both in the imaging data and in the physical space. We use 4 [mm] CT-Spots from Beekley Corp (Bristol, CT, USA), coated with infrared retro reflective material (Fig. 1(E)). Following the approach of Wang et al.[10], we use an automatic segmentation based on binary thresholding and region growing, followed by a classification of the segmented region. The centroids of segmented regions are calculated intensity-weighted using the voxel intensities of the imaging data.

Finally, the correct point correspondences are established and the transformation ${}^{Target}H_{CT}$ from the CT coordinates into the tracking coordinates is computed. This is done by a distance-weighted graph matching approach [11] followed by a point based registration algorithm [12]. Thus the data in the CT coordinate system can be transformed to the overlay image coordinate system by ${}^{Overlay}H_{CT} = {}^{Overlay}H_{Target}\ {}^{Target}H_{CT}$, with ${}^{Overlay}H_{Target}$ from equation 1.

3 Navigation Modes

3.1 Standard Slice Based Navigation Interface

In standard slice based navigation systems information is presented on two dimensional monitors. The slices displayed are controlled by the pose of the instrument. This pose, as well as the virtual extension of the instrument is drawn onto the slice. We implemented this in our visualization software to have all navigation modes presented in the head mounted display. Therefore, we project the slices at a fixed location in space (Fig. 2(d)).

3.2 Augmented Reality Visualization Modes

We implemented various augmented reality visualization modes. The requirement for navigation is the guidance of surgical instruments to a specific target point based on three dimensional imaging data. The requirement for the clinical integration is that no interaction is required to prepare the data (e.g. interactive segmentation or planning). All three implemented augmented reality navigation modes work directly on the DICOM data with the imaging data registered as described in section 2.2.

Volume Rendering. Multiple planes parallel to the image plane are clipped against the volume boundaries and rendered by interpolating within the volume and blending appropriately. Intensity values in the volume domain are mapped to the three dimensional color space using transfer functions in order to accentuate anatomically interesting structures. Additionally to the volume the extension of the instrument axis is visualized to provide navigational feedback (Fig. 2(a)).

Aligned Slice View. Any arbitrary slice can be visualized in-situ onto the surgical object. We visualize the plane defined by the target point and the sagittal, frontal, or axial direction as normal vector of the plane. The extension of the instrument axis is visualized. Its intersection with the plane is explicitly highlighted (Fig. 2(b)).

Instrument Aligned Orthoslice View. An orthoslice view is rendered spatially aligned along the instrument axis. The slices are controlled by the pose of the instrument. The intersection of the two planes corresponds to the instrument axis and along this line the user can see the extension of the instrument inside the anatomy (Fig. 2(c)).

(a) Volume rendering. (b) Aligned slice view. (c) Instrument aligned orthoslice view. (d) Standard slice based navigation.

Fig. 2. Different navigation modes that are rendered into the head mounted display

3.3 Hybrid Navigation Interface

The hybrid navigation interface combines the standard navigation interface and the above described augmented reality in-situ visualization modes into a single three dimensional interface (Fig. 3). The standard navigation is mapped just beside the real surgical object. Visualizing it close to the surgical object makes it visible at the same time as the in-situ modes project their information directly onto the surgical object. The advantage of the in-situ visualization i.e. intuitive and accurate for lateral dimensions, complements the advantages of the standard slice based navigation system i.e. high accuracy in all dimensions.

4 Experiments

For the evaluation and comparison of different combinations of navigation modes, we designed a phantom that mimics epidermal and osseous structures. We implanted 4 $[mm]$ metal spheres in a block of wood at a depth of approximately 30 $[mm]$. The surface of the phantom was covered with a silicone rubber compound which has properties similar to human skin. Additionally combined CT

(a) Navigation and volume rendering. (b) With aligned slice view. (c) With instrument aligned orthoslice view.

Fig. 3. Different hybrid navigation modes combining in-situ and standard slice based navigation

and infrared retro-reflective markers were attached to the phantom in order to allow image-to-physical object registration.

The fiducial registration error (FRE) using automatic marker segmentation and point based registration was 0.28 [mm]. We estimated a maximum target registration error (TRE) of 0.52 [mm] and a mean error of 0.43 [mm]$\pm$0.03 [mm] using the estimation approach of Fitzpatrick et al. [13].

Using the head mounted display, one of the metal spheres is highlighted in the in-situ view. The subject is then asked to navigate a tracked surgical drill to the target point using either one of the visualization modes (Fig. 2) or the proposed hybrid navigation user interface (Fig. 3).

During the experiment the time used to reach the target point and the distance between the tip of the drill and the centroid of the marked metal sphere is measured. The experiment was conducted by three trauma surgeons with different levels of experience, who marked 4 targets per visualization method, resulting in 28 measurements per subject.

5 Results

The mean accuracy of the three trauma surgeons for each representation is summarize in table 1.

Table 1. Comparison of different navigation modes in terms of accuracy and time required to complete the drilling task. The different modes were standard slice based navigation (NAV), volume rendering (VOLREN), aligned slice view (ASV), instrument aligned orthoslice view (IOV), and combinations of in-situ modes with the standard navigation.

	Surgeon A		Surgeon B		Surgeon C	
Navigation	error [mm]	time [s]	error [mm]	time [s]	error [mm]	time [s]
NAV	1.7 ± 1.1	128 ± 32	1.6 ± 0.7	140 ± 151	1.2 ± 0.7	35 ± 9
VOLREN	2.1 ± 1.1	134 ± 18	1.0 ± 0.4	91 ± 88	2.5 ± 1.6	24 ± 5
ASV	2.6 ± 0.8	76 ± 18	0.7 ± 0.3	50 ± 21	1.6 ± 0.9	49 ± 29
IOV	2.3 ± 0.9	98 ± 24	1.7 ± 0.6	57 ± 52	3.1 ± 2.9	47 ± 28
VOLREN + NAV	1.9 ± 1.5	106 ± 52	1.5 ± 1.3	52 ± 5	2.4 ± 0.5	26 ± 6
ASV + NAV	1.5 ± 0.2	83 ± 18	1.1 ± 0.6	50 ± 20	2.1 ± 1.5	24 ± 10
IOV + NAV	1.6 ± 0.6	95 ± 28	1.9 ± 0.5	84 ± 17	1.9 ± 0.2	26 ± 12

The results in terms of accuracy and speed of execution for each surgeon, as well as the distribution within the different visualization approaches lead to the assumption that the overall performance depends on the level of experience.

Surgeon A is a relatively inexperienced trauma surgeon, who does not use navigation systems in clinical routine. He performed the task with higher accuracy using the standard navigation interface in comparison to in-situ visualization modes. On the other hand he required more time with standard navigation than using in-situ visualization. Using the hybrid system, he achieved the same accuracy as with the standard navigation system but was significantly faster. This

shows the advantage of a hybrid mode over a standard slice based navigation in terms of speed and over augmented reality modes in terms of accuracy. The gain in speed can be related to the intuitive usability of the in-situ component especially for inexperienced surgeons.

Surgeon B is a very experienced trauma surgeon, who uses standard navigation systems regularly in the OR. Table 1 shows no significant difference in the accuracy throughout all visualization modes. However, more time was needed using the standard navigation mode. This results from difficulties in finding the right entry point and drill orientation. Here the hybrid modes seem to compensate for this effect. In addition he reported after the experiment that using the hybrid modes he had the confidence of the standard navigation mode.

Surgeon C is an experienced surgeon who is familiar with standard navigation systems and augmented reality visualization. Thus, he performed the task throughout fast with no significant difference. The relatively low accuracy compared to the other candidates was due to a miscalibration of the tooltip by ≈ 1 $[mm]$ in the direction of the drill axis. Since he was used to augmented reality he hardly used the standard slice based navigation in the hybrid mode.

6 Discussion

We presented different visualization modes for navigation in one single three dimensional user interface. Within the experiment, it was shown that the hybrid mode was improving the performance of a surgical action in accuracy compared to in-situ visualization mode and in terms of speed compared to standard navigation. This is only valid for performing the surgical action not for the overall time of a navigated procedure.

In addition to the measured results, the surgeons confirmed that the use of both, standard slice based navigation in combination with augmented reality visualization, can be of great benefit for surgery. The standard navigation system is not the most intuitive navigation interface. Problems especially arose with the position and orientation of the instrument in the lateral dimensions. In-situ visualization has its limitations in precise navigation, especially in depth.

The optimal placement of the standard navigation interface in space and the most useful mode of in-situ visualization depend on the clinical application. They will be defined as we will apply the system to real anatomy. With the hybrid navigation mode, we propose an alternative visualization interface that provides useful guidance for orthopedic and trauma surgery. It however needs to be totally integrated in the overall clinical workflow.

Acknowledgment. Special thanks to Frank Sauer, Ali Khamene, and Sebastian Vogt from Siemens Corporate Research for the design and implementation of the in-situ visualization system RAMP.

References

1. King, A.P., Edwards, P.J., Maurer, Jr., C.R., de Cunha, D.A., Hawkes, D.J., Hill, D.L.G., Gaston, R.P., Fenlon, M.R., Strong, A.J., Chandler, C.L., Richards, A., Gleeson, M.J.: A system for microscope-assisted guided interventions. IEEE Trans. Med. Imag. **19**(11) (2000) 1082–1093
2. Birkfellner, W., Figl, M., Huber, K., Watzinger, F., Wanschitz, F., Hummel, J., Hanel, R., Greimel, W., Homolka, P., Ewers, R., Bergmann, H.: A head-mounted operating binocular for augmented reality visualization in medicinedesign and initial evaluation. IEEE Trans. Med. Imag. **21**(8) (2002) 991–997
3. Sauer, F., Khamene, A., Bascle, B., Rubino, G.J.: A head-mounted display system for augmented reality image guidance: Towards clinical evaluation for imri-guided neurosurgery. In: Proc. of MICCAI, Springer-Verlag (2001) 707–716
4. 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 imageguided needle biopsy: Initial results in a swine model. Radiology **238**(2) (2006) 497–504
5. Azar, F.S., Perrin, N., Khamene, A., Vogt, S., Sauer, F.: User performance analysis of different image-based navigation systems for needle placement procedures. In: Proc. of the SPIE, Volume 5367, pp. 110-121 (2004). (2004) 110–121
6. Sauer, F., Wenzel, F., Vogt, S., Tao, Y., Genc, Y., Bani-Hashemi, A.: Augmented workspace: designing an ar testbed. In: Proc. of IEEE and ACM ISAR. (2000) 47–53
7. Hoff, W.A., Vincent, T.L.: Analysis of head pose accuracy in augmented reality. IEEE Trans. Visualization and Computer Graphics **6** (2000)
8. Sauer, F., Schoepf, U.J., Khamene, A., Vogt, S., Das, M., Silverman, S.G.: Augmented reality system for ct-guided interventions: System description and initial phantom trials. In: Medical Imaging: Visualization, Image-Guided Procedures, and Display. (2003)
9. Vogt, S., Khamene, A., Sauer, F., Niemann, H.: Single camera tracking of marker clusters: Multiparameter cluster optimization and experimental verification. In: Proc. of IEEE and ACM ISMAR. (2002) 127–136
10. Wang, M.Y., Maurer, Jr., C.R., Fitzpatrick, J.M., Maciunas, R.J.: An automatic technique for finding and localizing externally attached markers in ct and mr volume images of the head. IEEE Trans. Biomed. Eng. **43**(6) (1996) 627–637
11. Gold, S., Rangarajan, A.: A graduated assignment algorithm for graph matching. IEEE Trans. Pattern Anal. Mach. Intell. **18**(4) (1996) 377388
12. Umeyama, S.: Least-squares estimation of transformation parameters between two point patterns. IEEE Trans. Pattern Anal. Mach. Intell. **13**(4) (1991) 376–380
13. Fitzpatrick, J.M., West, J.B., Maurer, Jr., C.R.: Predicting error in rigid-body point-based registration. IEEE Trans. Med. Imag. **14**(5) (1998) 694–702

Virtual Fly-Over: A New Visualization Technique for Virtual Colonoscopy

M. Sabry Hassouna[1], A.A. Farag[1], and Robert Falk[2]

[1] Computer Vision & Image Processing Laboratory (CVIP), University of Louisville, Louisville, Kentucky, 40292
{msabry, farag}@cvip.uofl.edu
[2] Medical imaging, Jewish Hospital, Louisville, Kentucky
robert.falk@jhhs.org

Abstract. In this paper, we propose a new visualization technique for virtual colonoscopy (VC). The proposed method is called *Virtual Fly-Over*, which splits the entire colon anatomy into exactly two halves. Then, it assigns a virtual camera to each half to perform fly-over navigation, which has several advantages over both traditional fly-through and related methods. First, by controlling the elevation of the camera, there is no restriction on its field of view (FOV) angle (e.g., $> 90^o$) to maximize visualized surface areas, and hence no perspective distortion. Second, the camera viewing volume is perpendicular to each colon half, so potential polyps that are hidden behind haustral folds are easily found. Finally, because the orientation of the splitting surface is controllable, the navigation can be repeated at a different split orientation to overcome the problem of having a polyp that is divided between the two halves of the colon. Quantitative experimental results on 15 clinical datasets have shown that the average surface visibility coverage is $99.59 \pm 0.2\%$.

1 Introduction

Colorectal colon cancer is the third most common form of cancer and the second leading cause of death among cancers in the western world [1]. Since colorectal cancer is largely preventable, the colonoscopy screening test is recommended for all people age 50 and over. Although colonoscopy detects more than 90% of colorectal cancers, it is invasive, uncomfortable, inconvenient, and sometimes can not reach the colon caecum, resulting in an incomplete exam [2]. On the contrary, virtual colonoscopy (VC) is a computer-based alternative to real colonoscopy. VC is not intended to replace real colonoscopy, but rather to complement it by providing additional supportive information such as visualizing in both directions, passing high grade stenoses, and planning for surgery. In addition, it is the only alternative offered to those patients that are not comfortable with real colonoscopy or are severely ill [3].

The common visualization technique that tries to simulate the real colonoscopy is the virtual fly-through navigation [4, 5, 6], where a virtual camera with a specific FOV moves along a special planned path inside the colon to render its internal views. The direction of navigation either starts from the colon rectum side (antegrade), or from its caecum side (retrograde). In general, fly-through based-methods suffer from the following limitations: (1) the camera's FOV is limited, and hence results in a lower surface

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 381–388, 2006.

visibility coverage. (2) the navigation must be done in both antegrade and retrograde directions to maximize visualized surface areas, and hence it is very time consuming. Several visualization techniques have been proposed to overcome those shortcomings, which can be categorized as colon flattening and panoramic methods. The main idea behind colon flattening methods [7, 8, 9] is to initially transform the colon into a cylinder-like shape to reduce its overall tortuosity, and then to map the straightened colon onto a single image, which can be inspected from a single view point. Because flattening methods are based on geometric mapping, geometric distortion is inevitable. As a consequence, important diagnostic features can be altered or even eliminated.

Early work on panoramic methods was presented in [7]. The lateral side views of the colon have been captured by six views of 60^0 FOV each, which are then stitched together to form one view. The method has been extended in [10], where each centerline point is considered the center of a virtual cube. A virtual camera is located at that center to project the colon walls to each face of the cube using a 90^o FOV. The cube is then unfolded into a single plane. One of the main drawbacks of this method is that a polyp may split among several faces of the cube. In addition, the layout of the unfolded cube is hard to follow, since a physician must observe six views simultaneously in two different directions (horizontal and vertical). Such drawbacks are partially corrected in [11]. The back face of the cube is removed, while the rest are mapped into a single plane. The front face is located at the middle of the plane, while its nearest neighbor square faces are mapped to trapezoidal windows. Due to this nonlinear mapping, the method introduces geometric distortion to 80% of the cube surface area. The common drawback of all panoramic methods is that polyps behind haustral folds can be easily overlooked under the following scenarios: (1) A polyp is located along the camera optical axis, and hence its depth perception is lost by a physician. (2) A polyp is completely blocked by a haustral fold as shown in Figure 1(a).

In this paper, we present the *virtual fly-over* as a new visualization technique for virtual colonoscopy, that overcomes the limitations of related methods.

2 Methods

The proposed method consists of two steps. In the first step, the colon is split along its centerline into exactly two halves, while in the second step, each half is assigned a virtual camera that performs virtual fly-over navigation as shown in Figure 1(b). In order to split the colon into exactly two halves, we have to use a polygonal clipping surface that passes through the colon's centerline. Clipping a 3D surface by a polygonal surface is a very tedious task. A better methodology is to approximate the polygonal surface by the union of finite size planes. Unfortunately, this simplification does not even work, because clipping must be accomplished by implicit surfaces such as a sphere, cube, cylinder, or an infinite plane. To overcome this problem, we divide the colon surface into consecutive strips of triangles (rings). Each ring is split into two halves by an infinite plane that passes along its local centerline. Finally, each set of consecutive half rings are concatenated altogether to form one half of the colon. Initially the colon lumen is segmented using a seeded region growing algorithm. The centerline of the colon has been extracted using our recent work [4], which accurately and automatically computes

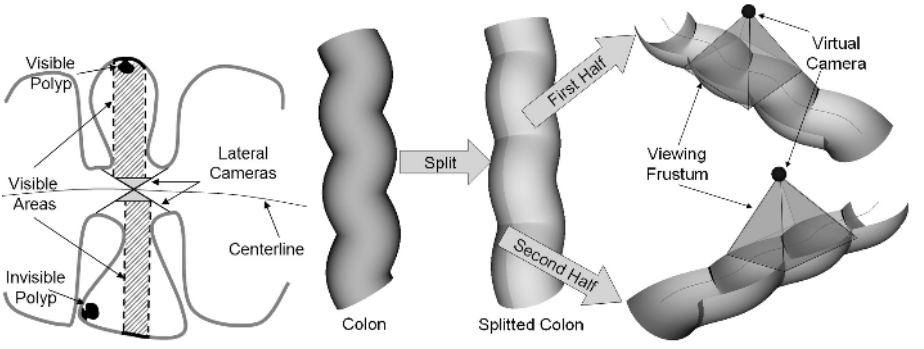


Fig. 1. (a) A schematic showing a cross section in a colon, where the lateral views of panoramic methods [10, 11] overlook a polyp that is completely blocked by a haustral fold. (b) Illustration of the proposed fly-over technique. Initially, the colon is split into two halves and then each half is assigned a virtual camera to perform fly-over navigation.

the centerline of non-tubular structures. Finally, the colon's surface is reconstructed in 3D using the marching cubes algorithm [12].

1. **Generation of Rings**
   Let C be a smoothed version of the colon's centerline, which is sampled uniformly to N_p points as shown in Figure 2(a). Each point p_i, where $1 \leq i \leq N_p$ is associated with a tangential vector $\overrightarrow{t_i}$ and a normal vector $\overrightarrow{n_i}$. The centerline C is also divided into N_s segments l_j, where $1 \leq j \leq N_s$. Each segment l_j is $1cm$ in length (i.e, 20 voxels for $0.5mm$ voxel size). The points of each segment l_j are assigned a specific label based on the following $label$ function.

$$\forall\, p_i \ \in l_j, \quad label(p_i) = j \tag{1}$$

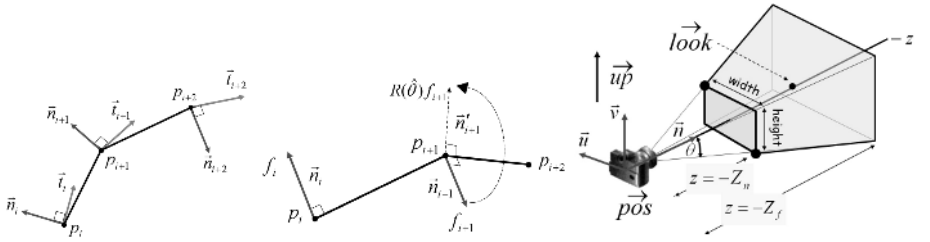
   In order to divide the colon surface into consecutive rings, we use a surface skinning method, where each mesh triangle is assigned the label of the nearest centerline segment l_j. As a consequence, the set of surface triangles of the same label form a ring. Let c_k be the geometric center of each surface triangle r_k, where $1 \leq k \leq N_r$ and N_r is the total number of surface triangles. Then, the skinning process is governed by the following equation.

$$\hat{p_i} = \arg\min_{p_i \in C} \|c_k - p_i\|^2, \quad label(r_k) = label(\hat{p_i}) \tag{2}$$

   Surface skinning is a fast process of computational complexity $O(N_r N_p)$. In Figure 3(a), we show the generated rings of a clinical colon dataset.

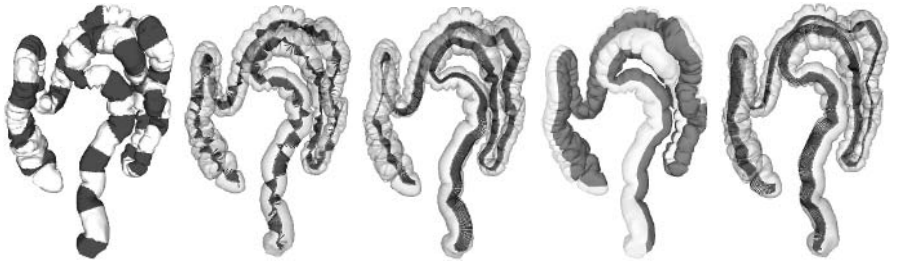
2. **Splitting Rings**
   Each ring R_j spans multiple centerline points, whose starting and ending points are q_j and q_{j+1}, respectively. In order to split R_j into two halves, we clip it using an infinite plane π_j, whose center is q_j and whose normal vector is given by

$$\overrightarrow{u_j} = (q_{j+1} - q_j) \times \overrightarrow{n_j} \tag{3}$$



(a) (b) (c)

Fig. 2. (a) The centerline C is uniformly sampled to N_p points. (b) Each point f_{i+1} is rotated around the line segment that is given by the points p_{i+1} and p_{i+2} until the rotation angle between the normal vectors $\boldsymbol{n}_{i+1}$ and $\boldsymbol{n}_i$ is minimized. (c) The pinhole virtual camera model.



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

Fig. 3. (a) Colon surface is divided into consecutive rings ($l_j > 1cm$). (b) The computed normals along the centerline may change their directions abruptly. (c) Aligned normals provide a good approximation of the clipping surface. (d) Splitting colon surface into two halves. (e) Changing the direction of the first normal vector results in a new split orientation surface.

To generate the entire clipping planes along all rings, we have to first compute the normal vectors along the centerline, which can be generated using a common method in computer graphics [13]. The normal vector $\overrightarrow{n_i}$ at each centerline point p_i is given by,

$$\overrightarrow{w_i} = \overrightarrow{up} \times \overrightarrow{t_i}, \quad \overrightarrow{n_i} = \overrightarrow{t_i} \times \overrightarrow{w_i} \tag{4}$$

where, $\overrightarrow{up}$ is an arbitrary vector, which is adaptively chosen at each point to be any of the Cartesian basis vectors as long as it is not parallel to $\overrightarrow{t_i}$. Since we approximate the polygonal splitting surface by the union of finite size planes, then each two consecutive planes must share an edge, which implies that each two consecutive normal vectors $\overrightarrow{n_i}$ and $\overrightarrow{n_{i+1}}$ are in the same plane. Unfortunately, due to the high tortuosity nature of the colon and hence its centerline, the generated normal vectors along the centerline using Eq. (4) may change their directions abruptly several times as shown in Figure 3(b). To overcome this problem, we propose the following method. Let $\overrightarrow{n_i}$ be described by the end points p_i and f_i, while $\overrightarrow{n_{i+1}}$ by p_{i+1} and f_{i+1} as shown in Figure 2(b). Each point f_{i+1} is rotated around the line segment that is given by the points p_{i+1} and p_{i+2} until the rotation angle between $\overrightarrow{n_{i+1}}$ and $\overrightarrow{n_i}$ is minimized, which can be formulated mathematically by the following optimization problem,

$$\hat{\theta} = \arg\min_{\theta} \|f_i - R(\theta) f_{i+1}\|^2 \tag{5}$$

where $R(\theta)$ is the rotation matrix. In Figure 3(c), we show the aligned normal vectors of the same colon dataset of Figure 3(b), while in Figure 3(d), we show the results of splitting the same dataset into exactly two halves.

3. **Controlling Split Orientation**
Since each normal vector $\overrightarrow{n_{i+1}}$ is aligned with respect to the previous vector $\overrightarrow{n_i}$, then by changing the direction of the first normal vector $\overrightarrow{n_1}$, the rest of the normal vectors will change their orientations as well. Therefore, $\overrightarrow{n_1}$ controls the orientation of the split surface. The first normal vector $\overrightarrow{n_1}$ of Figure 3(c) has been rotated by 90^o to yield a new split orientation surface as shown in Figure 3(e).

4. **Virtual Fly-Over**
Each colon half is assigned a perspective virtual camera, whose model is shown in Figure 2(c). The part of the pyramid that is enclosed by the clipping planes Z_n and Z_f is called the viewing volume or frustum. The orientation of each camera is controlled by four parameters that are computed automatically during fly-over navigation. The parameters are the camera position vector $\overrightarrow{pos}$, the look-at vector $\overrightarrow{look}$, which describes the direction of projection, the view-up vector $\overrightarrow{up}$, and the field of view angle θ. The look-at vectors of the first half ($l = 1$) and second half ($l = 2$) are given by,

$$\overrightarrow{look}_{i,l=1} = \overrightarrow{t_i} \times \overrightarrow{n_i}, \quad \overrightarrow{look}_{i,l=2} = -\overrightarrow{look}_{i,l=1} \tag{6}$$

The camera position and view-up vectors of each half are given as follows,

$$\overrightarrow{pos}_{i,l} = p_i + h_i \frac{\overrightarrow{look}_{i,l}}{\|\overrightarrow{look}_{i,l}\|}, \quad \overrightarrow{up}_{i,l} = \overrightarrow{t_i} \tag{7}$$

where h_i is the camera elevation from each centerline point p_i. According to Figure 2(c), the camera's FOV is given by,

$$\tan\left(\frac{\theta}{2}\right) = \frac{height}{2Z_n} \tag{8}$$

During centerline extraction, each point p_i encodes its distance from the colon's boundary $D(p_i)$. By making p_i the focal point of the camera, where the image plane is located, then the elevation of the camera h_i at each point p_i that maximizes surface visibility coverage, while maintaining the same angle θ is given by,

$$h_i = \frac{D(p_i)}{\tan(\theta/2)} \tag{9}$$

According to Eq. (9), there is no limitation on the FOV angle that maximizes surface visibility coverage. For example, for $\theta = 90^o$, $h_i = D(p_i)$, while for $\theta = 60^o$, $h_i = 1.73D(p_i)$.

3 Quantitative Validation

To quantitatively validate the proposed fly-over navigation method in terms of effectiveness (surface visibility coverage), clinical datasets of 15 patients were acquired using Siemens Sensation CT scanner. The dataset volume is $512 \times 512 \times 580$ with voxel size $0.74 \times 0.74 \times 0.75mm^3$.

In Table 1, we compare the surface visibility coverage of the proposed method with the traditional fly-through method for three clinical datasets, while navigating in the antegrade, retrograde, and both directions at 90^o FOV.

Table 1. Comparison of surface visibility by the fly-over and fly-through methods

Navigation Direction	Navigation Technique	Surface Visibility (%)		
		Dataset A	Dataset B	Dataset C
Antegrade	Fly-through	79.29	78.95	79.38
	Fly-over	**99.35**	**99.68**	**99.33**
Retrograde	Fly-through	77.89	75.74	79.00
	Fly-over	**99.6**	**99.91**	**99.68**
Antegrade + Retrograde	Fly-through	93.64	93.12	93.38
	Fly-over	**99.7**	**99.95**	**99.73**

4 Discussion and Conclusion

The validation experiment 15 clinical datasets have shown that the surface visibility coverage by the proposed method irrespective of the navigation direction is on average 99.59%$\pm$0.2. To ensure 100% surface coverage, invisible areas are clustered using a connected component algorithm and then properly visualized. Although panoramic methods have reported nearly the same surface visibility coverage [10, 11], they still suffer from other limitations as mentioned earlier. The increase in surface visibility coverage by the proposed fly-over method, while navigating in both directions is insignificant, which was expected because the camera viewing volume is perpendicular to each colon half. Therefore, fly-over navigation in either direction (antegrade or retrograde) is sufficient for nearly complete surface coverage. In Figure 4, we superimpose on one clinical dataset undetected surface patches (dark color) by both the fly-through and the proposed method. Notice that fly-over method greatly reduces blind areas.

In the first row of Figure 5, we show the rendered views during traditional fly-through navigation, while in the middle and last rows, we show the rendered views during fly-over navigation of the first and second colon halves, respectively. While some rendered views of both fly-through and fly-over are showing the same number of polyps as shown in the first two columns of Figure 5, some rendered views by the proposed fly-over method are showing more polyps than the corresponding fly-through views as shown in the last two columns of the same figure. The evaluation time was on average 26 minutes by fly-through method in both directions and 13 minutes (including 3 minutes for surface reconstruction, skinning, splitting, and rendering) by fly-over method on a 3Ghz Pentium computer with 2 GB memory and nVidia Quadro FX 3000 graphics card.

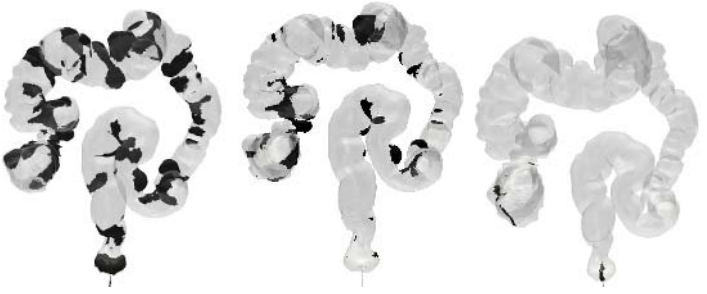


Fig. 4. Comparison of surface visibility coverage by (a) Fly-through in antegrade direction, (b) Fly-through in both directions, (c) Fly-over in antegrade direction.

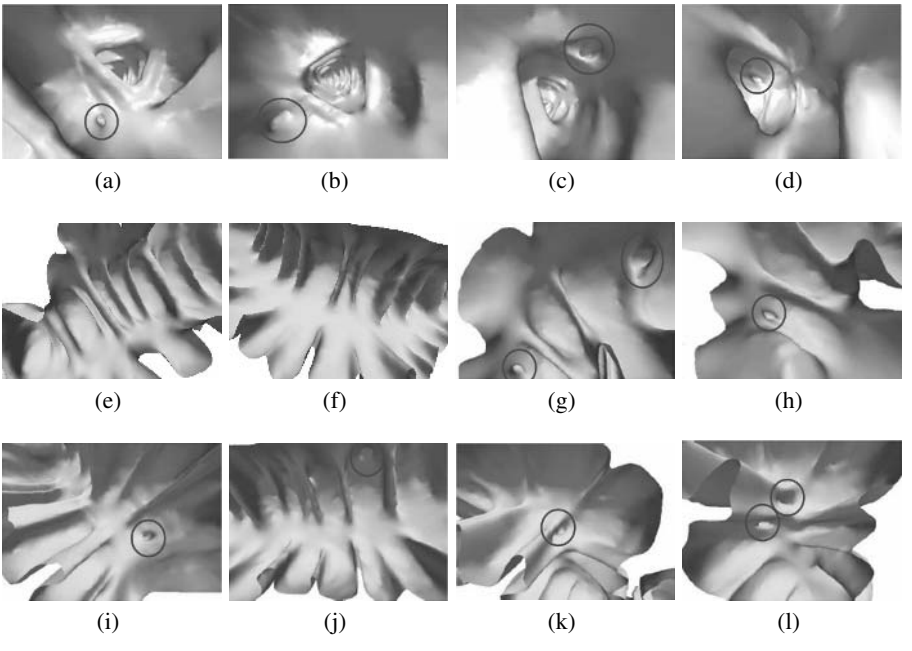


Fig. 5. Colon rendered views. (a-d) Fly-through navigation. (e-h) First colon half using fly-over. (i-l) Second colon half using fly-over.

To conclude, in this paper we have presented a new visualization technique for virtual colonoscopy, whose advantages can be enumerated as follows. Due to the nature of the fly-over navigation, the viewing volume is perpendicular to each colon half, which results in higher surface visibility coverage and higher sensitivity. By controlling the elevation of the camera, there is no restriction on its FOV angle to maximize visualized surface area, and hence the proposed method does not suffer from perspective distortion. Because the split orientation is controllable, the navigation can be repeated at a different split orientation if a polyp is divided between the halves of a colon. Follow-up

research by our group includes large scale patient screening study to assess the proposed approach versus related methods, and to design a CAD system for polyp detection.

Acknowledgment. The colonoscopy dataset used in this paper were provided by 3DR Inc., Louisville, Kentucky.

References

1. Abbruzzese, J., Pollock, R.: Gastrointestinal Cancer. 1 edn. Springer (2004)
2. Macari, M., Berman, P., Dicker, M., Milano, A., Megibow, A.: Usefulness of ct colonography in patients with incomplete colonoscopy. J. Roentgenol **173** (1999) 561–564
3. Baert, A.L., Sartor, K.: Virtual Endoscopy and Related 3D Techniques. Springer (2001)
4. Hassouna, M.S., Farag, A., Falk, R.: Differential fly-throughs (dft): A general framework for computing flight paths. In: MICCAI, Palm Springs, California, October 26-29 (2005)
5. Bouix, S., Siddiqi, K., Tannenbaum, A.: Flux driven fly throughs. In: Computer Vision and Pattern Recognition. (2003) 449–454
6. Kang, D.G., Ra, J.B.: A new path planning algorithm for maximizing visibility in computed tomography colonography. **24**(8) (2005) 957 – 968
7. Paik, D., Beaulieu, C., et al.: Visualization modes for ct colonography using cylindrical and planar map projections. Journal of Computer Tomography **24**(2) (2000) 179–188
8. Haker, S., Angenent, S., Tannenbaum, A., Kikinis, R.: Nondistorting flattening maps and the 3d visualization of colon ct images. **19**(7) (2000) 665–671
9. Oda, M., Kitasaka, T., Hayashi, Y., Mori, K., Suenaga, Y., ichiro Toriwaki, J.: Development of a navigation-based cad system for colon. In: MICCAI. (2005) 696–703
10. Vos, F.M., van, R.E., Serlie, I.W.O., et al.: Three-dimensional display modes for ct colonography: conventional 3d virtual colonoscopy versus unfolded cube projection. Radiology **228** (2003) 878–885
11. Geiger, B., Chefd'hotel, C., Sudarsky, S.: Panoramic views for virtual endoscopy. In: MICCAI. (2005) 662–669
12. Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3d surface construction algorithm. In: In Proc. of SIGGRAPH. (1987) 163–169
13. Hill, F.J.: Computer Graphics Using OpenGL. Prentice Hall PTR (2000)

Viscoelasticity Modeling of the Prostate Region Using Vibro-elastography

S.E. Salcudean[1], Daniel French[1], S. Bachmann[1], R. Zahiri-Azar[1], X. Wen[1], and W.J. Morris[2]

[1] Department of Electrical and Computer Engineering, University of British Columbia, Canada
tims@ece.ubc.ca
[2] Radiation Oncology, British Columbia Cancer Agency, Canada

Abstract. We present an ultrasound vibro-elastography system designed to acquire viscoelastic properties of the prostate and peri-prostatic tissue. An excitation stage imparts low-frequency (<20 Hz), limited amplitude ($< \pm$ 2mm), broadband vibratory motion to an endorectal transducer, along a radial/transversal direction. The induced tissue motion is estimated from ultrasound radio-frequency data and is used to estimate the mechanical frequency response of tissue to the excitation at different spatial locations. This can be used to determine the spatial distribution of various mechanical parameters of tissue, such as stiffness and viscosity. Phantom and *in-vivo* images are presented. The results obtained demonstrate high phantom and tissue linearity and high signal-to-noise ratio.

1 Introduction

Brachytherapy is a widely used treatment for prostate cancer. The treatment consists of the placement of radioactive seeds in the prostate using trans-perineal insertion of needles, in a pattern that delivers a sufficient radiation dose to kill the cancer while maintaining a tolerable dose to the urethra and rectum [1]. Biochemical and clinical relapse rates are very low with current brachytherapy procedures. However, improvements are still required to reduce the rate of complications, which include erectile dysfunction, urinary retention, incontinence and rectal injuries. Complications have been associated with high dose loads to large portions of the prostate, with high urethral dose and with needle trauma. Complication rates may be reduced by more accurate needle insertions.

Seed placement errors are caused by poor visualization of the prostate and by prostate movement due to needle forces. B-mode ultrasound (US) and X-ray fluoroscopy are typically used for image guidance during brachytherapy. Prostate segmentation in US images is unreliable, whether done manually or automatically [2], especially at the apex where the prostate blends into the muscles of the pelvic floor, and at the base, where the anterior portion is iso-echoic with the neck of the bladder. During each needle insertion, needle forces move the prostate [3] and cause seed placement errors [4]. Prostate swelling and deformation caused by edema, and motion due to respiration also generate errors [5].

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 389–396, 2006.

Thus there is a need to reliably localize the prostate based on US images and to predict tissue deformation due to needle forces.

The use of tissue elastic models for needle insertion planning has been proposed before. In [6], rigid needle insertion optimization based on a 2-D linear Finite-Element Method (FEM) tissue model has been proposed. In [7], needle motion planning based on 2-D FEM tissue and flexible needle models is reported. Needle targeting using discrete spring models is described in [8]. All these methods require knowledge of the tissue elastic properties, which are also required for the development of medical simulators [9].

The elastic properties of the prostate have been measured before *ex-vivo* [10]. These can also be extracted from strain imaging systems, such as presented in [11] for prostate cancer detection and image-guided biopsy [12]. However, traditional strain imaging systems with quasi-static and/or manual excitation produce images that have significant variance and are operator dependent.

This paper presents a system for the acquisition of patient-specific elastic properties of tissue in the prostate region using a dynamic ultrasound elastography method called vibro-elastography (VE) [13]. The system produces stable, operator-independent images that can be used to estimate stiffness as well as other properties of tissue. The hardware, signal processing and software necessary for this purpose are presented in Section 2 of this paper. Phantom and *in-vivo* imaging results are presented in Section 3, followed by a discussion of the results in Section 4 and conclusions in Section 5.

2 Methods

In VE, tissue is compressed by a vibrating actuator driven by low-pass filtered white noise, typically up to 20Hz, while tissue is imaged with US over a short time interval (1-10 s) (Fig. 1). The US radio-frequency (RF) data is used to determine time sequences of tissue displacements $x_i(t)$ at spatial locations i, typically along the same US RF A-line. Assuming that the response of tissue to the excitation is linear, mechanical frequency responses or *transfer functions* (TFs) $H_i^j(\omega), \omega \in I\!R$ of the displacement at spatial location j with respect to the displacement at spatial location i viewed as an input, and the associated *coherence functions* $C_i^j(\omega)$, can be computed by well-known techniques [14]:

$$H_i^j(\omega) = P_{x_i x_j}(\omega)/P_{x_i x_i}(\omega); \quad C_i^j(\omega) = |P_{x_i x_j}(\omega)|^2/(P_{x_i x_i}(\omega)P_{x_j x_j}(\omega)) \ . \ (1)$$

where $P_{x_i x_j}(\omega)$ denotes the cross spectral density between $x_i(t)$ and $x_j(t)$. $C_i^j(\omega) \in [0,1]$ indicates the proportion of input energy at location i at ω that appears at the output j at the same ω and therefore it is an indicator of both system linearity and signal-to-noise ratio. One has high confidence in the TF estimate $H_i^j(\omega)$ if the coherence function $C_i^j(\omega)$ is close to 1 at ω.

Mechanical tissue properties such as stiffness and viscosity can be derived from these TFs. For example, $|H_i^0(0)|$ estimates the static deformation at location i due to a unit displacement at location 0. Therefore $|H_i^{i+1}(\omega)/H_{i-1}^i(\omega)|_{\omega=0}$ will

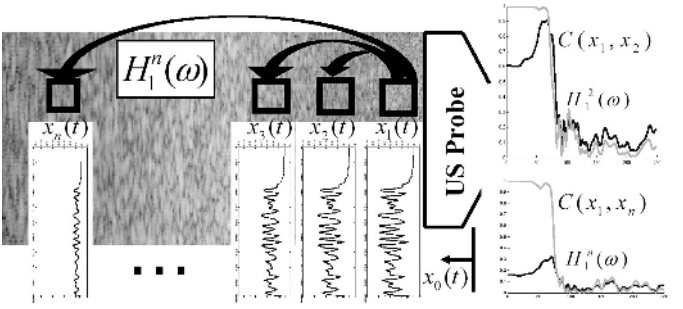


Fig. 1. Vibro-elastography with actuated US transducer, measured displacements, frequency responses and coherence functions

show relative tissue compliance between consecutive spatial locations. Stiffness estimation based on TFs relies on elastic behavior that can be averaged over an entire low frequency range and therefore it is superior to static elastograms [13]. Stiffness changes of 10% and stiff (2X) enclosures of 0.7 mm can be detected with a linear 8 MHz transducer [13]. A trans-rectal US (TRUS) VE system has been constructed. The system utilizes an Ultrasonix 500 RP US machine with a biplane TRUS transducer (Ultrasonix Medical Corp.). The transducer is mounted on an actuating system, described in Section 2.1, which is used to apply compression waves to the rectal wall. US images are acquired and processed on the Ultrasonix 500 RP PC-based machine using the signal processing techniques described below.

2.1 Vibro-elastography Actuation Stage Design

The TRUS VE actuation system can translate and orient the TRUS transducer and can impart to it radial vibratory motion (Fig. 2). It has three main parts: a vibrating mechanism, a rotational stage and a transverse stage. The maximum displacement of the endorectal probe from its nominal center can be set by limit stops in the range ±0-5mm, while experimental results have shown that ±2 mm is sufficient for the acquisition of prostate images. The actuation system is mounted on a standard brachytherapy stabilizer for coarse positioning.

Low inertia coreless motors (#118778 Maxon Precision Motors) with digital encoders (QPhase 3076A007-5000-0, Quantum Devices Inc), current amplifiers and a PCI interface card (LCAM and Q8, Quanser Inc.) have been used in the actuation system. A PC-based real-time system (WinCon by Quanser and Simulink from The MathWorks) and a user interface that allows desired waveforms to be executed have been implemented. The closed loop position control bandwidth achieved with the present controller is of 10 Hz for small amplitude motion, while a pre-compensator allows open loop position excitation at much higher frequencies. A number of safety features have been implemented and include system kill switches for two system operators, fused motors, software limit stops and procedures that enforce a properly initialized start-up sequence.

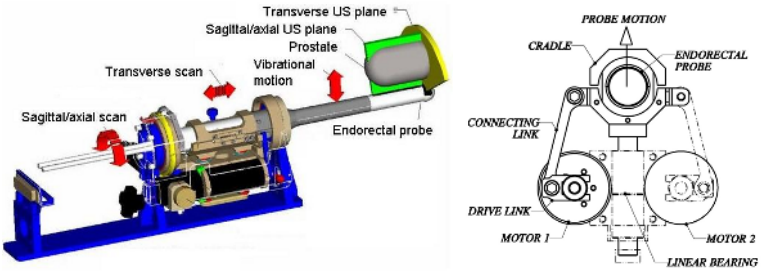


Fig. 2. VE actuation system schematic (left) and transversal schematic view of the vibratory system (right). The dual imaging probe (imaging planes shown through the prostate model) is mounted in a housing and is vibrated along its axial plane by two motor-driven four-bar linkages, along a direction set by the rotational actuator, with its insertion depth adjusted by the translational actuator. Volume images can be acquired using the transverse or the sagittal imaging planes.

2.2 Vibro-elastography Signal Processing

Real-time Motion Estimation. Tissue motion due to TRUS probe compression and relaxation is estimated by time domain cross-correlation. Each A-line of RF data is divided into N short, equal-length overlapping windows. The displacement $x_i(kT)$ of the i^{th} window from sample time kT to sample time $kT+T$ can be obtained by maximizing its cross-correlation to the RF A-line sampled at time $kT + T$. Because adjacent segments have similar displacements when the US image sampling period T is small, the search for cross-correlation maximization can be bracketed to a very small predicted interval [15]. Sub-sample resolution is achieved by interpolation (quadratic, cosine). The normalized cross-correlation is monitored. When it drops below a certain threshold (e.g., 0.9), a "recovery search" that samples a larger segment of the RF A-line, not just the small predicted interval, is performed. While we have not carried out an extensive evaluation during brachytherapy, in phantom studies and limited *in-vivo* studies, the recovery search is seldom triggered - in fewer than 2% of processed windows, allowing processing of windows of 16,000 pixels at 30 Hz.

Real-time Frequency Response Estimation. The displacements $\{x_i(kT)\}$ of the i^{th} A-line RF data window are buffered in segments of fixed length and overlap. Estimates of $H^i_M(\omega)$, $i = 1, ..., M-1, M+1, ..., N$ from a reference A-line RF window to each of the other windows in the A-line are obtained using Welch's periodogram method [16], in which the spectral densities are efficiently updated every time a new data segment is accumulated. The periodic updates are obtained by using all the data, by discarding old data or by discounting the effect of old data exponentially using a "forgetting factor". Discarding or discounting old data results in TF update based on recent data which enables the operator to make real-time adjustments to the VE imaging system.

Display of Images. VE generates a spatial distribution of frequency responses with a wide choice of parameters for display. In all images presented in this paper we display the L_2 norm of the differences between TFs at different spatial

locations $\|W(H_i^{i+1} - H_{i-1}^i)\|_2 = \|W(H_M^{i+1}/H_M^i - H_M^i/H_M^{i-1})\|_2$, where W is a frequency-dependent weight matching the input excitation frequency. In the low range of excitation frequencies considered in this paper, the acquired responses are fairly flat and therefore this norm approximates tissue compliance.

2.3 Phantom Construction

To tune and characterize the TRUS VE system, a prostate phantom has been constructed. It consists of a rubber balloon filled with 14% gelatin gel and 2% cellulose (by mass) (G9382-B gelatin, S-5504 cellulose, Sigma-Aldrich Inc.). The balloon is embedded in 10 cm×10 cm×10 cm substrate of 10% gelatin gel and 2% cellulose. The gel in the balloon is approximately twice as stiff as the substrate and has similar echogenic properties. A cylindrical hole in the phantom mimics the rectum and permits insertion of the endorectal ultrasound probe.

3 Results

In the images presented, the RF A-lines are split into 80 windows each with 50% window overlap. Axial images contain 128 RF A-lines along the 60 mm long transducer crystal. Transverse images contain 114 RF A-lines. Image depth is

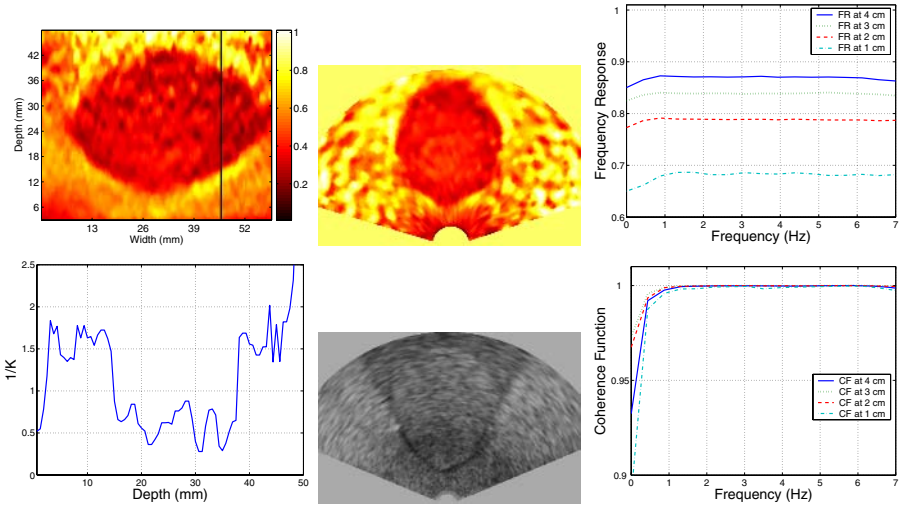


Fig. 3. Prostate phantom images. Sagittal VE image (top left), relative compliance along the marked VE A-line showing a change of a factor of 2 (bottom left - inclusion is stiffer), transversal images (top middle - VE, bottom middle - B-mode), TFs (top right) and corresponding coference functions (bottom right) along the marked A-line are displayed. The motion applied is 4 mm peak-to-peak, band-pass filtered (0.5-10Hz) white noise. Transfer functions are averaged over 10 seconds, and are computed with reference to the motion in the focal area of the firing crystal at a depth of 25mm. Colour coding is HOT as defined in Matlab - see index bar.

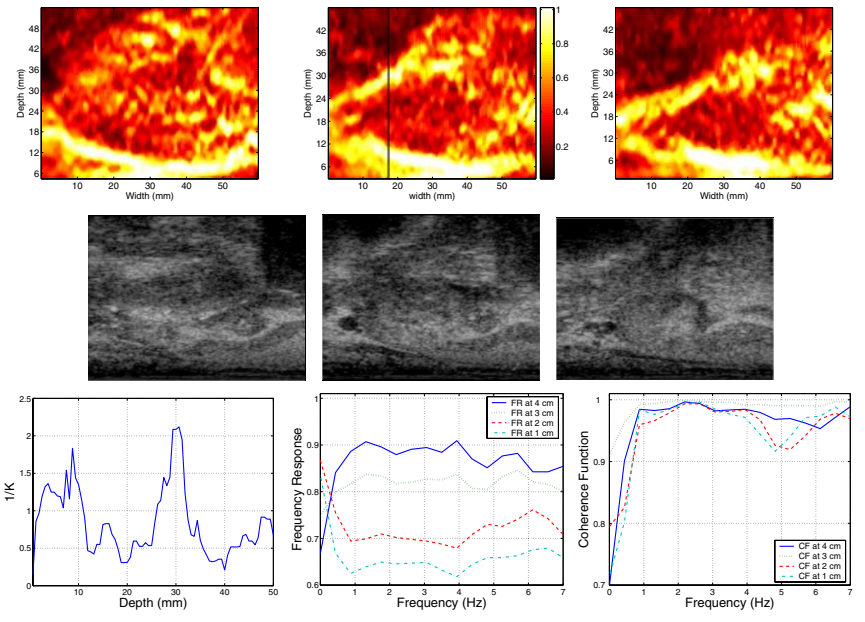


Fig. 4. Prostate sagittal and para-sagittal images near the apex. VE - top row, showing sagittal plane (at 90 degrees), and para-sagittal planes at 80 degrees and 70 degrees. The corresponding B-mode images are shown in the middle row. The bottom row shows the relative compliance (left), five TFs (middle) and the corresponding coherence functions (right) along the marked A-line. The motion applied was 2 mm peak-to-peak, band-pass filtered (0.5-4.5 Hz) white noise. TFs are computed with reference to a window at 25 mm depth.

50 mm in all cases, transducer frequency is 5 MHz, and sampling frequency is 40 MHz. The RF-data acquisition and processing frame rate was 14 Hz. TFs were displayed at 2 Hz, allowing real-time tuning of imaging parameters (e.g., compression, amplitude, and frequency content of the input).

Images of the phantom described above are shown in Fig. 3. A protocol for acquiring VE images from patients has been designed and images have been acquired from three patients in the operating room just prior to the brachytherapy procedure. Images are presented in Fig. 4 and Fig. 5. 1D median filtering with a kernel of 5 samples has been applied to displacement estimates, which show a normalized correlation coefficient above 0.9 in over 90% of windows. A 2D median filter with a 3x3 pixel kernel has been applied to VE images, which are also decay-compensated in a standard manner.

4 Discussion and Future Work

The VE images obtained, as shown in Figs. 3, 4, and 5, show very good delineation of the prostate gland by comparison to B-mode US images. Furthermore, they show significant and repeatable details, including stiff regions within the

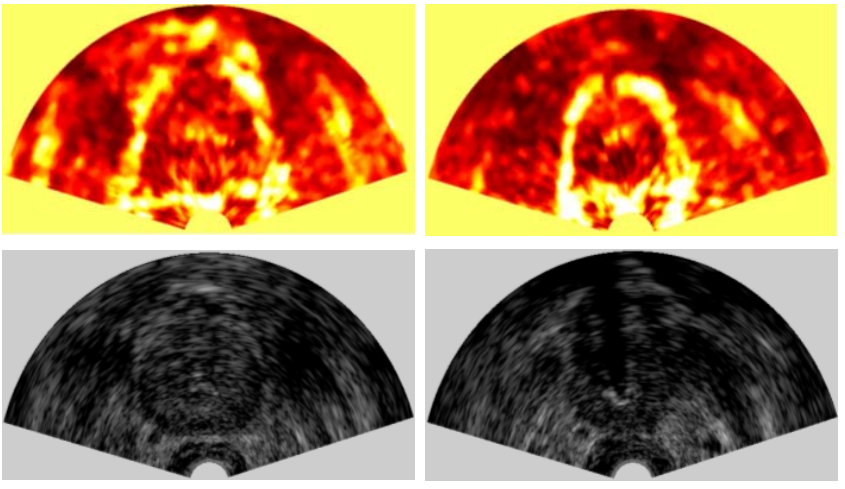


Fig. 5. Prostate transversal images. Top row shows VE images of the midgland (left) and apex (right), bottom row corresponding B-mode images. The motion applied was the same as in Fig. 4, TFs were computed with reference to a window at 25mm depth.

prostate. The results show surprisingly high coherence over the frequencies of excitation, over 0.9 in 80% of TFs shown in the images. While the excitation frequency is quite low, this indicates good confidence in the measurement.

For good estimation of the transfer function, there must be sufficient frequency content in the transducer motion which provides the tissue compression. In this work, the motion of a window (indexed by M) in the focal area of the A-line was used as a reference. While this motion may not be as "white" as the transducer motion which is under computer control, we have not synchronized yet the US transducer location to the RF data acquisition. Such synchronization may allow improved TF estimation.

The display of other TF parameters is possible. Most notably, in preliminary work, we were able to correlate the changes in TF phase with changes in viscosity. This may also provide image contrast. Unfortunately, because of sampling rate limitations, we have not collected data at high enough frequency to resolve phase lag. We plan to do this in the future.

In order to use the results presented in a needle insertion simulator suitable for brachytherapy treatment planning, a larger tissue volume will need to be imaged. We have been limited to the 50 mm depth in the present system due to hardware limitations of the US machine we are presently using. This limitation will dissappear as we upgrade the new Ultrasonix Sonix US machine.

5 Conclusion

We presented a system for the acquisition of patient specific visco-elasticity images of the prostate and peri-prostatic tissue using ultrasound vibro-elastography.

The advantages of the approach include the acquisition of a measurement confidence measure through the coherence function, the ability to tune the frequencies at which measurements are taken and the ability to acquire repeatable, operator independent images. While only a few patients have been imaged so far, the results are encouraging. The measured frequency responses can be used to delineate the prostate and in needle insertion models and planners useful in brachytherapy treatment. Further research is required to determine the potential usefulness of such a system in prostate cancer diagnosis.

References

1. Zietman, A.L.: Localized prostate cancer: Brachytherapy. Curr Treat Options Oncol. **3**(5) (2002) 429–36
2. Shao, F., Ling, K.V., Ng, W.S., Wu, R.Y.: Prostate boundary detection from ultrasonographic images. Journal of Ultrasound in Medicine **22** (2003) 605–623
3. Lagerburg, V., et al: Measurement of prostate rotation during insertion of needles for brachytherapy. Radiothera. Oncol. **77** (2005) 318–23
4. Alterovitz, R., et al.: Needle insertion and radioactive seed implantation in human tissues: Simulation and sensitivity analysis. In: Proc. IEEE Int. Conf. Rob. Aut. Volume 2. (2003) 1793–1799
5. Nag, S., et al: Intra-operative planning and evaluation of permanent prostate brachytherapy: Report of the American Brachytherapy Society. Int. J. Rad. Onc. Biol. Phys. **51**(5) (2001) 1422–1430
6. Alterovitz, R., et al.: Sensorless planning for medical needle insertion procedures. In: Proc. IEEE/RSJ IROS Conf. Volume 4. (2003) 3337–3343
7. DiMaio, S.P., Salcudean, S.E.: Interactive simulation of needle insertion models. IEEE Trans. Biomedical Engineering **52**(7) (2005) 1167–1179
8. Glozman, D., Shoham, M.: Flexible needle steering and optimal trajectory planning for percutaneous therapies. In: Proc. MICCAI. Volume 3217. (2004) 137–144
9. Goksel, O., et al: 3D needle-tissue interaction simulation for prostate brachytherapy. In: Proc. MICCAI. (2005) 827–834
10. Krouskop, T.A., Wheeler, T.M., Kallel, F., Garra, B.S., Hall, T.: Elastic moduli of breast and prostate tissues under compression. Ultrason. Imaging **20**(4) (1998) 260–274
11. Klauser, A., et al.: Real-time elastography for prostate cancer detection. J. Urology **171**(4 Suppl.) (2004) 477
12. Konig, K., et al.: Initial experiences with real-time elastography guided biopsies of the prostate. The Journal of Urology **174** (2005) 115–117
13. Turgay, E., Salcudean, S., Rohling, R.: Identifying mechanical properties of tissue by ultrasound. Ultrasound in Medicine and Biology **32**(2) (2006) 221–235
14. Bendat, J.S., Piersol, A.G.: Random Data: Analysis and Measurement Procedures, 2nd Edition. Wiley Interscience. (1986)
15. Zahiri-Azar, R., Salcudean, S.E.: Time domain cross correlation with prior estimates. In: 3^{rd} Int. Conf. US. Meas. Imag. Tissue Elasticity, Cumbria, UK (2004)
16. Oppenheim, A., Schafer, R.: Digital Signal Processing. Englewood Cliffs, NJ, Prentice-Hall. (1975)

Simultaneous Reconstruction of Tissue Attenuation and Radioactivity Maps in SPECT

Yi Tian[1], Huafeng Liu[1], and Pengcheng Shi[2,3]

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

Abstract. The importance of accurate attenuation correction in single photon emission computed tomography (SPECT) has been widely recognized. In this paper, we propose a novel scheme of simultaneous reconstruction of the tissue attenuation map and the radioactivity distribution from SPECT emission sinograms, which is obviously beneficial when the transmission data is missing for cost or efficiency reasons. Our strategy combines the SPECT image formation and data measurement models, whereas the attenuation parameters are treated as random variables with known prior statistics. After converting the models to state space representation, the extended Kalman filtering procedures are adopted to linearize the equations and to provide the joint estimates in an approximate optimal sense. Experiments have been performed on synthetic data and real scanning data to illustrate abilities and benefits of the method.

1 Introduction

Single photon emission computed tomography provides functional information of the physiological processes of interests. Recently, SEPCT has enjoyed increasingly widespread use, because its costs and technology are within the reach of most clinical nuclear medicine facilities. The physical basis of SPECT is to reconstruct the radioisotope labeled tracer distribution externally through collection of photons that are emitted from inside the human body. However, due to the attenuation of the photon during it traveling through various tissues, the image quality and the quantitative accuracy of SPECT reconstructions degrade greatly from the true distribution. Thus, there have been considerable efforts within the SPECT community aimed at incorporating attenuation correction into any given reconstruction method [5].

Earlier approaches of attenuation correction in SPECT largely fall into two groups. The first one attempts to measure the attenuation map based on transmission scanning, including an external radionuclide source [4,5], CT [7], or segmented MRI data [8]. However, because the emitted photon traverses only part of the patient's anatomy before reaching detector, the problem of photon attenuation in SPECT has proven to be very difficult. There is still a need to

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 397–404, 2006.

develop appropriate strategies for accurate and reliable attenuation map recovery based on transmission scanning data. The second group attempts to recover the attenuation map without adding a separate transmission scan. Some approaches involve using a known/segmented object contour with a priori distribution of the attenuation coefficient. However, while manual contour delineation is very complex and time consuming, automatic edge-detection remains one of the open issues in the medical image analysis. On the other hand, in practical situations, especially for pathological data sets, it is almost impossible to have the *exact* patient-dependent attenuation information *a priori*. In addition to the prior methods, the incorporation of consistency conditions and statistical modeling into simultaneous estimation of activity and attenuation map from emission data has received a large amount of attention [10,1,6]. These iterative reconstruction algorithms require careful modeling of the imaging system response model which is subject to a number of physical effects. How to process the effects of the uncertainties in the model remains an important issue [3,9].

In this paper, we present a joint estimation approach to recover the tissue attenuation and radioactivity distributions simultaneously from emission sinograms. Our efforts have two significant novel aspects. First, with the SPECT imaging and measurement processes in state space representation, the attenuation parameters are treated as random variables with known prior statistics in our effort. Thus, the proposed scheme unified the attenuation map reconstruction problem from transmission scan or emission sinogram data into a generalized framework. Secondly, while in general accurate identification of attenuation from emission data is difficult, our approach undertakes the data and model uncertainties naturally by formulating the problem as an extended state estimation effort on the attenuation parameters. In our current implementation, the extended Kalman filter (EKF) procedures are adopted to linearize the augmented state representation and provide joint/simultaneous estimates in the minimum-mean-square-error sense. An evaluation study on simulated phantom data and real imaging data is described, where experimental results have shown great promise for such strategy.

2 Methodology

2.1 SPECT Scan Modeling

In SPECT imaging, the emission sinogram data, acquiring by a rotating detector head around the patient at each angel, can be represented lexicographically as a column vector $y = \{y_i | i = 1, ..., M\}$. Here i indicates different projection defined by rotating angel and different detector bin, and M is the total number of projections. During an emission scan, photon will be attenuated by body tissue before reaching the detector along its travelling path. Assuming that the activity distribution and attenuation map are given by column vectors $x = \{x_j | j = 1, ..., N\}$ and $\mu = \{\mu_j | j = 1, ..., N\}$, one photon's survival probability after attenuation is equal to linear integration along its path in the direction

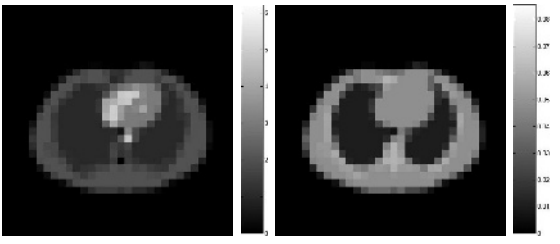


Fig. 1. A digital Zubal phantom: activity distribution (*left*) and attenuation coefficient distribution (*right*)

towards the detector. Where x_j and μ_j represent the the activity and the linear attenuation coefficient in voxel j, and N is the total number of image voxels.

Let l_{ijk} represent the probability that the photon generated in voxel j passing voxel k and attenuated by μ_k along a projection line i when emitted into the detector. The photon survival probability considering about attenuation can be represented by a matrix A with elements $[A]_{ij} = a_{ij}$ as

$$a_{ij} = \mathbf{exp}\left(-\sum_k l_{ijk}\mu_k\right) \tag{1}$$

If we use d_{ij} to indicate the probability that a photon produced in voxel j reaches the front face of detector along projection i without attenuated, then represent all of these probability with a 3 dimensional vector $[L]_{ijk} = l_{ijk}$ and a system matrix $[D]_{ij} = d_{ij}$, the mathematical model of the SPECT emission measurement could be written as [1]

$$y = (A \cdot D)x + noise = \left(e^{-[L\mu]} \cdot D\right)x + noise \tag{2}$$

or in a discrete form

$$y_i = \sum_j \left[\mathbf{exp}\left(-\sum_k l_{ijk}\mu_k\right) d_{ij}x_j\right] + noise \tag{3}$$

2.2 State Space Representation of SPECT Imaging

In emission tomography such as SPECT, the goal is to reconstruct the emission/radioactivity distribution x from the projected measurement data y (the photon counts). The state equation of the imaging system, which describes the radioactivity of the pixels, can be written in the form of

$$x(t+1) = Sx(t) + v_s \tag{4}$$

with some initial activity x_0. The system noise v_s models the statistical uncertainty of the imaging model. In general, equation (4) represents the dynamic

[1] Note that the symbol '$\cdot$' means the dot product of 2 matrixes.

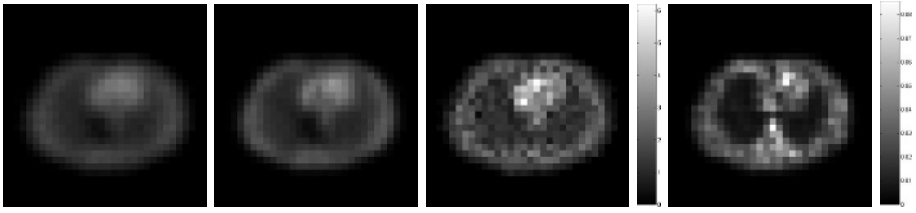


Fig. 2. From left to right: activity maps recovered by FBP, EM-ML, and our method, and the attenuation map estimated by our framework

changes of the state variable x, and it reduces to the conventional static reconstruction problem when the transition matrix S is an identity matrix. The associated measurement equation (2), which describes the observations provided by the imaging data $y(t)$, can be expressed in a more general form:

$$y(t) = Gx(t) + e_o \tag{5}$$

where G is a known measurement matrix, and e_o is the *measurement noise*.

2.3 Augmented State Space Representation of SPECT Imaging

Equations (4) and (5) provide a so called state-space representation of the SPECT system with assumed known attenuation coefficients. In practice, however, the attenuation data are not precisely known. In this work, with the attenuation coefficients treated as random variables with known or assumed prior distributions, we are trying to determine the best estimates of the spatial distributions of the activity and attenuation map simultaneously.

In order to perform the joint estimation, the *unknown* state vector x is augmented by the *unknown* attenuation coefficient vector μ to form the new state vector $z = [x\ \mu]^T$. Since μ is a temporally constant, spatially varying parameter and treat x as static for one frame of the image, equation (4) is converted to the new augmented state equation:

$$z(t+1) = z(t) + v(t) \tag{6}$$

Here v represents the uncertainties of the model, with $E[v(m)v(n)] = Q(t)\delta_{mn}$.

In a similar fashion, the new augmented measurement equation is derived from equation (2) and (5):

$$y(t) = f(z(t)) + e(t) \tag{7}$$

with

$$f(z(t)) = \left(e^{-[L\mu(t)]} \cdot D\right) x(t) \tag{8}$$

$$E[e(t)] = 0, E[e(m)e(n)'] = R(t)\delta_{mn} \tag{9}$$

where e represents the uncertainties of the measurement. By solving $\hat{z}(t)$ from above nonlinear state space equation, the estimation of activity map $\hat{x}(t)$ and attenuation map $\hat{\mu}(t)$ can be obtained simultaneously.

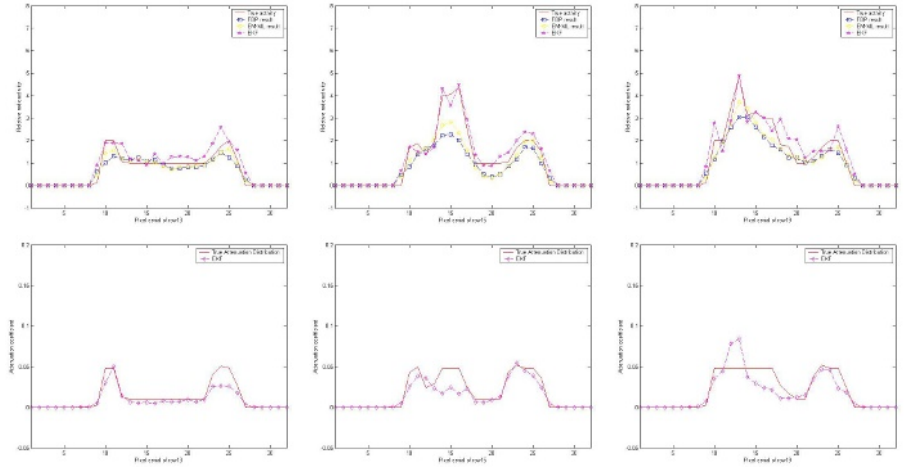


Fig. 3. Top: Vertical profiles along 13^{th} column (*left*), 15^{th} column (*middle*), 19^{th} column (*right*) of recovered activity maps. Bottom: Vertical profiles along 13^{th} column (*left*), 15^{th} column (*middle*), 19^{th} column (*right*) of recovered attenuation maps.

It is known that the noises in emission sinogram is Poisson distributed. However, by applying the Anscombe transformation, the Poisson noise could be converted into Gaussian one [2]. If x is Poisson distributed with mean equal to $\bar{x}$, then $(x + 3/8)^{1/2}$ can be approximated as Gaussian distributed with its mean equal to $\bar{x} + 1/8$ and its variance approximately equal to 0.25 [2]. Thus this form of state space formulation (equations (6) and (7)) leads to a solution of the filtering problem using the extended Kalman filter (EKF).

2.4 Joint Estimation of Tissue Attenuation and Radioactivity Maps: An EKF Solution

To solve the nonlinear state space equations (6)-(7) with a measurement description (8), we adopt an EKF approach to estimate the state variables. In order to apply the EKF, the nonlinear function f is recursively linearized around the most recent estimate of $z(t)$. By taking gradient of the transformation function at time t according to $H(t) = \frac{\partial f(z)}{\partial z^T}|_{z=\hat{z}^-(t)}$, the estimated value $\hat{z}(t)$ of vector $z(t)$ can be carried out by the following prediction-correction iterative process.

Prediction

$$\hat{z}^-(t+1) = \hat{z}(t) \tag{10}$$

$$P^-(t+1) = P(t) + Q \tag{11}$$

Correction

$$\hat{z}(t) = z^-(t) + K(t)[y(t) - f(\hat{z}^-(t))] \tag{12}$$

$$P(t) = [I - K(t)H(t)]P^-(t) \tag{13}$$

$$K(t) = P^-(t)H^T(t)[H(t)P^-(t)H^T(t) + R]^{-1} \tag{14}$$

Here, the Kalman gain matrix $K(t)$ drives the improvement of $z\hat{(}t)$ and P in the correction step by weighting the new information coming from the next system measurement $y(t)$ and the previously stored information in the prediction step. The initialization at time $t = 0$ of the state vector $z(0)$ and the relevant covariance matrix $P(0) = E\{[z(0) - \hat{z}(0)][z(0) - \hat{z}(0)]^T\}$ are given at the beginning of the iteration. The covariance of the measurement error R and system error Q is assumed to be known and set to time-invariant.

3 Experiments and Discussions

3.1 Digital Phantom Study

A digital Zubal thorax phantom with known radioactivity concentrations and attenuation coefficient distribution is used, as shown in Fig.1, to evaluate our method. The resolution of the original image is 32 by 32 pixels, and 1024 projections over 360 degrees are simulated for a rotating SPECT detector head. To generate realistic data, we simulate the single photon emission events considering tissue attenuation and detection probabilities, the measured sinogram y is created based on equation (2), here the noise mainly is 30 percents scatter events.

With this set of synthetic data, three types of reconstruction methods (FBP, EM-ML and our method) are then used. The activity maps obtained by our method and other methods are shown in Fig.2, and the attenuation map reconstructed from the emission data by our simultaneous estimation is also shown in the same figure. The EM-ML and FBP method with unknown attenuation map settings produce some noticeable errors. The proposed joint estimation framework gives the best estimates. From the vertical profiles, as shown in Fig. 3, we can draw the same conclusion more clearly.

3.2 Real Phantom Scanning Data with a Clinical Modality

The second data set used in this study was acquired on the SIEMENS ECam^{duet} SPECT scanner with a real cylinder phantom. The dimension of the phantom is 200mm (diameter) x 290mm (depth). A Teflon rod and two two hollow PMMA cylinders with diameters of $50mm$ are inserted in the phantom's volume, as shown in Fig. 4.

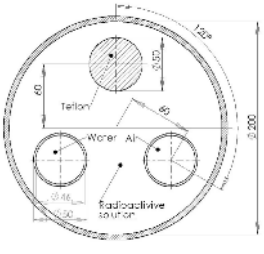


Fig. 4. A physical imaging phantom with 3 different material rods inside

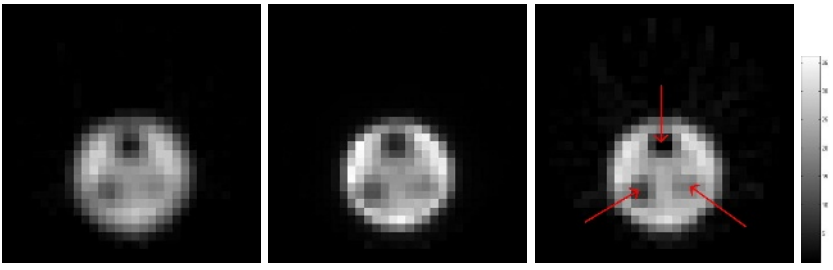


Fig. 5. Reconstructed activity map of real phantom data by FBP (*left*), EM-ML (*middle*), and our method (*right*), where the arrows indicate cold areas

The phantom is filled with $^{99m}\mathbf{Tc}$ concentration with a total radioactivity of 20mCi (100kBq/cc) and the two hollow cylinder rods are filled with air and pure water respectively. The phantom was scanned with a SIEMENS ECam^{duet} ECT scanner by two detector head rotating at total 64 angle position around 180 degree and the acquiring time at each position is 30 seconds. The final sinogram data has 64×128 projections for each slice. Then FBP, EM-ML method and our algorithms as described in the previous sections have been applied to recover images from the measured data, as shown in Fig.5. It is obvious that our method produces the best reconstruction results, especially for the three cold areas.

3.3 Discussions

Attenuation correction from transmission data is a well-documented research topic. Actually, the attenuation and activity distributions can be simultaneously reconstructed based on our proposed framework if both transmission and emission data are available. In this case, the measurement vector y consists of the emission data y_1 and transmission scanning measurement y_2: $y(t) = [y_1\ y_2]^T$. Thus, the measurement equation of the state space representation is converted into: $y(t) = \begin{bmatrix} G(t) & 0 \\ 0 & A_{tran} \cdot D \end{bmatrix} \begin{bmatrix} x(t) \\ x_{tran}(t) \end{bmatrix} + e(t)$. Where the matrix G is the dot product of A, D and L which defined in section 2.1 as $G(t) = A \cdot D = e^{-[L\mu(t)]} \cdot D$, where $x_{tran}(t)$ is the known effective activity distribution of the external transmission source and the diagonal matrix A_{tran} is the transmission attenuation factor matrix along every transmission projection, which diagonal element has the form as $a_{ii}^{tran} = \mathbf{exp}\left(-\sum_j l_{ij}\mu_j t\right)$. Here l_{ij} represents the effective intersection length along projection i passing voxel j, and μ_j is the attenuation coefficient of voxel j. This form of formulation leads to a solution of the filtering problem based on extended Kalman filter algorithm.

As a straightforward extension, based on nonlinear $\mathcal{H}_\infty$ filter, we can reconstruct the activity map with taking into account for the attenuation effects. The mini-max $\mathcal{H}_\infty$ criterion does not impose restrictions on the noise type, as long as it is of finite energy, and thus is more robust and less sensitive to noise variations and modeling assumptions. Detailed investigations on these issues are underway.

Acknowledgments. This work is supported by the National Basic Research Program of China (2003CB716100), the Hong Kong Research Grants Council (HKUST6151/03E), the NSF of China for Innovative Research Groups (60021201), and the Open Foundation of State Key Laboratory of Modern Optical Instrumentation at Zhejiang University (LMOI-0403).

References

1. A.Krol, J.E.Bowsher, S.H.Manglos, D.H.Feiglin, M.P.Tomai, and F.D.Thomas. An em algorithm for estimating spect emission and transmission parameters from emissions data only. *IEEE Trans. Med. Imag.*, 20:218–232, 2001.
2. F.J.Anscombe. The transformation of poisson, binomial and negative-binomial data. *Biometrika*, 35:246–254, 1948.
3. F.J.Beekman, C.Kamphuisa, M.A.Kingb, P.P.van Rijka, and M.A.Viergevera. Improvement of image resolution and quantitative accuracy in clinical single photon emission computed tomography. *Comput. Med. Imaging Graph*, 25:105–111, 2001.
4. H.H.Barrett and W.Swindell. *Radiological Imaging: The Theory of Image Formation, Detection, and Processing.* CA: Academic, San Diego, 1981.
5. H.Zaidi and B.Hasegawa. Determination of the attenuation map in emission tomography. *J. Nucl. Med.*, 44:291–315, 2003.
6. J.Nuyts, P.Dupont, S.Stroobants, R.Benninck, L.Mortelmans, and P.Suetens. Simultaneous maximum a posteriori reconstruction of attenuation and activity distributions from emission sinograms. *IEEE Trans. Med. Imag.*, 18:393–403, 1999.
7. J.S.Fleming. A technique for using ct images in attenuation correction and quantification in spect. *Nucl. Med. Commun.*, 10:83–97, 1989.
8. N.P.Rowell, J.Glaholm, M.A.Flower, B.Cronin, and V.R.McCready. Anatomically derived attenuation coefficients for use in quantitative single photon emission tomography studies of the thorax. *Eur. J. Nucl. Med.*, 19:36–40, 1992.
9. R.M.Lewitt and S.Matej. Overview of methods for image reconstruction from projections in emission computed tomography. *Proc. of the IEEE*, 91:1588–1611, 2003.
10. Y.Censor, D.Gustafson, A.Lent, and H.Tuy. A new approach to the emission computerized tomography problem: simultaneous calculation of attenuation and activity coefficients. *IEEE Trans. Nucl. Sci.*, 26:2275–2279, 1979.

Statistical Finite Element Model for Bone Shape and Biomechanical Properties

Laura Belenguer Querol[1], Philippe Büchler[1], Daniel Rueckert[2], Lutz P. Nolte[1], and Miguel Á. González Ballester[1]

[1] MEM Research Center, University of Bern, Switzerland
{laura.belenguer, philippe.buechler, lutz.nolte, miguel.gonzalez} @memcenter.unibe.ch
[2] Visual Information Processing, Department of Computing, Imperial College London, UK
dr@doc.ic.ac.uk

Abstract. We present a framework for statistical finite element analysis combining shape and material properties, and allowing performing statistical statements of biomechanical performance across a given population. In this paper, we focus on the design of orthopaedic implants that fit a maximum percentage of the target population, both in terms of geometry and biomechanical stability. CT scans of the bone under consideration are registered non-rigidly to obtain correspondences in position and intensity between them. A statistical model of shape and intensity (bone density) is computed by means of principal component analysis. Afterwards, finite element analysis (FEA) is performed to analyse the biomechanical performance of the bones. Realistic forces are applied on the bones and the resulting displacement and bone stress distribution are calculated. The mechanical behaviour of different PCA bone instances is compared.

1 Introduction

Current orthopaedic implant design techniques make use of limited information about the target bone, such as a set of shape parameters (e.g. lengths and angles), which fail to capture the complexity of real anatomical shapes. Furthermore, evaluation and optimization of the implant is done by manual fitting and fixation procedures, applied on a small set of cadaver bones in a trial-and-error process to find the optimal shape and mechanic properties. Thus, it is difficult to assess whether the set of implant shapes will fit most of the population. The use of poorly fitting implants may lead to bone wear and the need for revision surgery to replace the implant. The availability of a statistical model of bone shape and biomechanical properties across a target population would allow to virtually test the implants on a representative set of bones, generated by sampling the statistical model. Thus, it would be possible to optimise implant shape as to fit a maximum of the target population.

This work proposes a methodology to perform statistical finite element analysis. We first build a statistical model combining bone shape and density information, via principal component analysis of CT images (in which Hounsfield units are

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 405 – 411, 2006.

proportional to bone density). This model will contain the average bone and the principal modes of variation, spanning a probability density function in a combined shape and intensity space, which allows generating further valid instances of bones. Afterwards, finite element analysis (FEA) is performed by simulating a set of forces acting on the bone, thus obtaining a statistical distribution of the stress (or other biomechanical response measures) at all points within the bone. This allows performing statistical statements of biomechanical performance across a given population.

Section 2 describes the methodology for construction of the combined statistical model of shape and intensity. Finite element analysis based on this model, and in particular its use for statistical assessment of bone stress under particular force loading conditions, are explained in Section 3. Results are reported in section 4. Finally, discussion and directions for future work are provided.

2 Statistical Shape and Intensity Model

Prior to model construction, the original image data (in our case CT images) are segmented to isolate the structure under consideration. Based on this segmentation, a mask is applied so only points within the region of interest are considered for further processing.

Correspondences are established between shapes by non-rigid registration. One of the images is taken to be the reference image, which will be deformed non-rigidly to the remaining instances. In order to rule out biases induced by the choice of the registration algorithm, we have tested two different registration methods, ITK [10] and Pasha [2], finding the results almost identical (Fig. 1).

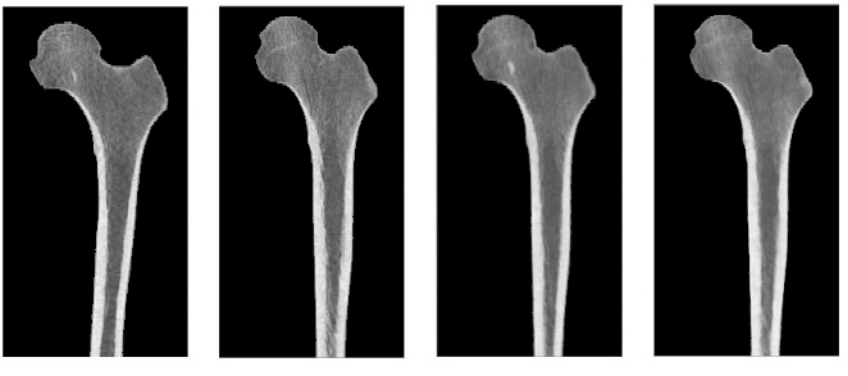


Fig. 1. Non-rigid registration examples using Pasha and ITK. From left to right: Reference and target masked images; Non-rigid registration with Pasha; Non-rigid registration with ITK.

In a second step, based on the deformation fields obtained from the registration process, we build vectors of corresponding positions and image intensities. The reference image can be described as in [8]:

$$c_{REF} = (x_1, y_1, z_1, I_1, ..., x_n, y_n, z_n, I_n)^T \quad (1)$$

where n is the number of voxels in the region of interest and I_i is the intensity at voxel (x_i, y_i, z_i). Similarly, each of the other images can be described as a vector of the same length:

$$c_j = (x_1 + \Delta x_1^j, y_1 + \Delta y_1^j, z_1 + \Delta z_1^j, I_1^j, ... \\ x_n + \Delta x_n^j, y_n + \Delta y_n^j, z_n + \Delta z_n^j, I_n^j)^T \quad (2)$$

where $(\Delta x_i^j, \Delta y_i^j, \Delta z_i^j)$ is the displacement vector at position (x_i, y_i, z_i), and I_i^j is the intensity of the voxel $(x_i + \Delta x_i^j, y_i + \Delta y_i^j, z_i + \Delta z_i^j)$ in image j.

The method used to compute a statistical description of the shape and intensity vectors is principal component analyis (PCA). We thus, obtain the average vector of the positions and intensities, $\bar{c}$, and the principal modes of variation, as the columns of the matrix ϕ, computed by singular value decomposition of the covariance matrix. For compactness, the number of these vectors can be reduced by setting a threshold in the variance to be explained by our statistical model [4]. New instances of the model can be computed as linear combinations of the modes of variation:

$$c = \bar{c} + \phi\, b \quad (3)$$

where b is a vector of scalar coefficients modulating the contribution of each mode of variation. This effectively generates new images depicting valid bone shapes and corresponding bone densities.

3 Finite Element Analysis

Finite Element (FE) analysis is a numerical technique to solve partial differential equations over domains of complex shapes. FE techniques find a natural application in biomechanical studies, such as is the case for structural analysis of orthopaedic implants. FE models are useful to assess the design, position and fixation of new implants [1].

A technique to generate FE models representing the target population in terms of shape and mechanical properties is proposed in this study. The statistical models obtained in the former step, average and modes of variation are used to create 3D bone solids representing instances of the population. The obtained bone geometries are then meshed with a finite number (about 3500) of 3D tetrahedrons. Four FE nodes, one for each corner, define the shape of the elements.

The mechanical properties used in the model are inhomogeneous and depend on the bone density distribution. Since calibrated CT scans were used for the construction of the statistical model, instances of the model will maintain a proportional relationship between the bone relative density and the grey level in the

images. It has been shown that the bone's Young's modulus can be obtained directly from the bone density [3, 9]:

$$E = \begin{cases} 60 + 900\rho^2, & \rho < 0.46\left(g/cm^3\right) \\ 2875\,\rho^3, & otherwise \end{cases} \tag{4}$$

where E is the Young's modulus in MPa and ρ is the bone relative density (g/cm^3).

The Poisson ratio is chosen equal to 0.3, because this parameter is not dependent on bone density. The loading conditions correspond to a 1600 N force (2 times body weight) on the humeral head while the diaphyseal part of the bone is maintained fixed.

The displacement and bone stress distribution corresponding to the defined loading condition is calculated, and they lead to define bone stiffness, as the ratio between the force and the displacement, $k = F / d$. The average and variation of the stiffness over the population is obtained.

4 Results

The combined statistical model was computed from 11 CT proximal femur data sets, which were cropped to only contain the region of the femur (ROI dimensions 175×150×200, voxel size 1×1×1.65 mm^3). The femur was segmented and other structures were masked as to not affect the registration. As mentioned in Section 2, one of the CT scans was used as the reference and the other images are registered to it. PCA is then applied to obtain the mean and modes of variation of combined shape and intensity. The top row of Fig.2 shows the mean and the effect of applying the first mode of variation (±2 standard deviations).

In a next step, we transfer the results obtained from the statistical model to create 3D bone solids, defined by a finite number of tetrahedrons (generated using Amira®, Mercury Computing Systems, USA). To obtain the mechanical properties we read the intensity values at the mesh node coordinates. We obtain their relative density and the Young's modulus as described in (4). The middle row of Fig. 2, shows the relative bone density for the bones generated from the instances of the statistical model when applying the first mode of variation. It can be observed that the larger variations in shape and density are focused on the cortical area.

The loading setup defined in Section 3 is implemented and solved using Patran (MSC Software, USA). The von Misses stress for each of the bone instances generated above is shown in the bottom row of Fig. 2. These meshes represent the variation in biomechanical properties as consequence of the modulation of the first principal mode of variation of the combined statistical model.

The computed global stiffness for the three bone instances under vertical loading are (from left to right: -2sd, mean and +2sd): 0.36 N/mm^2, 0.54 N/mm^2, 2.14 N/mm^2. It can be observed that the variation induced by the first mode of variation leads to a change in the angle of the femoral shaft and reduced bone thickness. This affects the overall stability of the bone, as can be seen by the increase in the stress of the cortical bone (Fig.2, bottom), and the sharp increase in stiffness for the +2sd case (right).

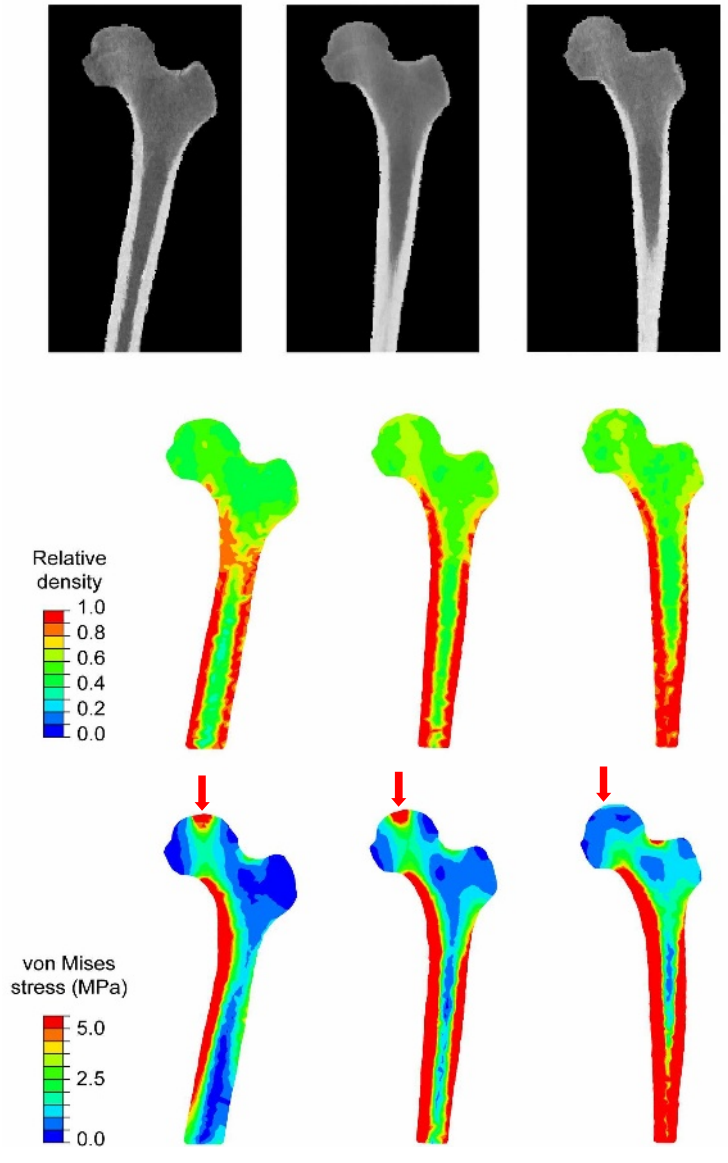


Fig. 2. Combined statistical model of shape and intensity and mechanical properties. From left to right: -2sd, average and +2sd. From top to bottom: combined statistical model of shape and intensity; relative density distribution; stress distribution corresponding to a vertical load of 1600N applied at the arrow locations.

5 Discussion and Conclusions

We presented a framework for statistical biomechanical assessment including a combined statistical model of shape and intensity, in combination with finite element

analysis. The application of our methods for modelling bone shape and mechanical behaviour has been shown.

The statistical model construction is based on non-rigid registration. This has the advantage that no landmarks or parametric representations need to be defined. By combining shape and intensity information into the same PCA evaluation, the model contains information about the correlation between shape and density. This could allow to predict bone density when only shape is known, and this is one of our directions of future work.

The FE analysis shown in this work could be improved by addressing a set of experiment setup issues. First, we employed tetrahedral meshes, which are known to have poor accuracy as compared to other element types. Further, the bone geometries were meshed independently from each other, resulting in different FE meshes for the different bones. This leads to difficulties in the direct comparison of the stress distribution between the different bone shapes under consideration, as well as discrepancies in the location of the loading forces applied. The alternative is to deform a pre-defined mesh, such as is done by [12]. However, it is extremely difficult to ensure the validity of the mesh for further FEA under arbitrary deformations.

The target application is orthopaedic implant design. Virtual testing of new implants will in the future replace cadaver testing. Further, being able to study the whole range of bone shapes and densities of the target population will lead to better fitting implants, as well as a considerable cost reduction in the design process. In order to assess the appropriateness of an implant, further development should be done to define the different scenarios of the implant, in terms of positions were the it is likely to be placed and the force loading conditions.

A complementary application of these techniques will be the patient-specific pre-clinical evaluation of an implant, taking into account the particular skeletal anatomy, bone quality, and implant position to assess the biomechanical performance of the implant on the patient.

Acknowledgements

We would like to thank Dr. Xavier Pennec from INRIA for sharing his non-rigid registration software, Pasha.

This research was funded by the Swiss National Science Foundation through its National Center of Competence in Research (NCCR) on Computer Aided and Image Guided Medical Interventions (http://co-me.ch).

References

1. Büchler, P., and Farron, A., Benefits of an Anatomical Reconstruction of the Humeral Head During Shoulder Arthroplasty: A Finite Element Analysis. Clinical Biomechanics, *19* (1):16-23, 2004
2. Cachier, P., Bardinet, E., Dormont, D., Pennec, X., and Ayache, N.: Iconic Feature Based Nonrigid Registration: The PASHA Algorithm. Computer Vision and Image Understanding. Special Issue on Nonrigid Registration, 89(2-3):272-298, February.-March, 2003.

3. Carter, D.R., and Hayes, W.C.: The compressive behavior of bone as a two-phased porous structure, Journal of Bone and Joint Surgery, 58A(7), pp. 954-962, 1977.
4. Cootes, T.F. and Taylor C.J.: Statistical models of appearance for medical image analysis and computer vision. Proc. SPIE Medical Imaging, 2001.
5. Cootes, T.F., Edwards, G.J. and Taylor, C.J.: Active Appearance Models. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.23, No.6, pp.681-685, 2001.
6. Cootes, T.F., Taylor C.J, Cooper, D.H and Graham, J.: Active shape models - Their training and application, Computer Vision Image Understanding, vol 61, no 1, pp 38–59, 1995.
7. Frangi, A.F., Rueckert, D., Schnabel, J.A., and Niessen, W.J.: Automatic construction of multiple-object three-dimensional statistical shape models: Application to cardiac modeling. IEEE Transactions on Medical Imaging, 21(9):1151-1166, 2002.
8. González Ballester, M.A., Pennec, X., Linguraru, M.G. and Ayache, N.: Generalised image models and their application as statistical models of images. Medical Image Analysis, 8(3) 361-9, 2004.
9. Rice, J.C., Cowin, S.C. and Bowman, J.A.: On the dependence of the elasticity and strength of cancelous bone on apparent density, Journal of Biomechanics, Vol. 21(2), pp. 155-168, 1988.
10. Rueckert, D., Frangi, A.F., and Schnabel, J.A.: Automatic construction of 3D statistical deformation models of the brain using non-rigid registration. IEEE Transactions on Medical Imaging, 22(8):1014-1025, 2003.
11. Rueckert, D., Sonoda, L.I., C. Hayes,. Hill, D.L.G., Leach, M.O., and Hawkes, D.J.: Non-rigid registration using free-form deformations: Application to breast MR images. IEEE Transactions on Medical Imaging, 18(8):712-721, 1999.
12. Yao, J. and Taylor,R.: Tetrahedral Mesh Modeling of Density Data for Anatomical Atlases and Intensity-Based Registration. Proc. MICCAI 2000. Pittsburgh, PA, USA: Springer, 2000.

Fetus Support Manipulator with Flexible Balloon-Based Stabilizer for Endoscopic Intrauterine Surgery

Hongen Liao[1], Hirokazu Suzuki[2], Kiyoshi Matsumiya[2], Ken Masamune[2], Takeyoshi Dohi[2], and Toshio Chiba[3]

[1] Graduate School of Engineering, the University of Tokyo
[2] Graduate School of Information Science and Technology, the University of Tokyo
[3] Department of Strategic Medicine, National Center for Child Health and Development
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
liao@miki.pe.u-tokyo.ac.jp

Abstract. A novel manipulator is described for stabilizing fetus and preventing it from free-floating during the endoscopic intrauterine surgery. Minimally invasive endoscopic fetal surgery enables intrauterine intervention with reduced risk to the mother and fetus. We designed and fabricated a prototype of a fetus supporting manipulator equipped with a flexible bending/curving mechanisms and a soft balloon-based stabilizer. The flexible bending and curving mechanisms enable the stabilizer to reach the target sites within the uterus under an ultrasound-guidance. The balloon-based stabilizer could be inserted into the uterus with a small incision for entry. The accuracy evaluation showed that the maximum error of the curving mechanism was as small as 7 mm and the standard deviation of the bending mechanism was just 1.6°. In the experiments using a fetus model, the manipulator could be well controlled under ultrasound guidance and its curving mechanism with the balloon-based stabilizer could be clearly visualized during the implementation of fetus model supporting. The manipulator has the potential to be used in minimally invasive intrauterine surgery, though further improvements and experiments remain to be carried out.

1 Background

Myelomeningocele is one of the most common congenital defects of the central nervous system [1]. It is a neural tube defect with spinal bonny malformation leading to an incomplete spinal canal. This can result in the dorsal protrusion of the spinal cord and its covering membranes. Myelonmeningocele accounts for about 75% of all cases of spina bifida and may occur in 1 per 1000 infants.

Since the neural damaging process like the obstacle of cerebrospinal fluid circulation, irrigation and absorption start before birth irreversibly, current neonatal surgical repair does not aim to reverse the central nervous damages, but aims merely to preserve the residual nervous function. Previous studies showed that the intrauterine environment may cause secondary injury to the spinal cord that is already dysplastic [2]. This suggested that in utero repair of the myelomeningocele could minimize this secondary injury and improve the life-long neurologic outcome [3].

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 412 – 419, 2006.

Surgical treatment of the myelomeningocele during 19-25 weeks of gestation is expected to improve the central/spinal cord nerve functions in utero preventing an association of hydrocephalus. The rationale of this procedure is based on the fact that the compliance of the fetal nervous system is high enough to be normalized in terms of its morphology due to the immature fetal myelin tissues not yet fully formed at this developmental stage. Furthermore, regeneration of damaged neural tissues is potentially better in earlier stages.

In recent years, endoscopic technology has provided a less invasive approach to surgical treatment. Minimally invasive endoscopic fetal surgery enables intrauterine intervention with reduced risks to the mother and fetus [4]. The problem is that the fetus is floating in the amniotic fluid and its movement will disturbe the surgical treatment. Then, it is necessary to develop a fetus-holding device to support the fetus within the uterus. In a previous research, we attempted to develop a stainless steel fetal stabilizer [5]. The outcome of this research was unfavorable although the fetus was fixed well. Other device was a suction type silicone tube stabilizer equipped with small holes to apply negative pressures on the fetal skin [6]. However, this device was likely to cause fetal tissue congestion during a long time surgery.

To overcome these issues, we propose a newly-developed manipulator with a flexible mechanism and a balloon-based supporting stabilizer. The outcome of mechanical performance tests and surgical practice tests using a fetal model under ultrasound guidance will be also described.

2 Methods

We have designed and fabricated a prototype of the fetus supporting manipulator, aimed at proposing a flexible and soft stabilizer for supporting the fetus in the endoscope intrauterine surgery. In this section, fetus supporting method, system requirements, configuration, and the apparatus of the prototype are described.

2.1 Requirements of Fetal Surgery

The gestational age selected for myelomeningocele treatment is at 19-25 weeks corresponding to the characteristics of the fetus. The fetus in the target gestation is very small with a length of about 30 cm and weight of about 500 g. The volume of the amniotic fluid is about 500 ml.

The fetal surgery is different from the other forms of surgery as follows. First, the fetus is fragile; the surgical instrument needs to be made small and flexible enough from hurting the fetus and the placenta. Second, the insertion sites for the surgical instrument into the uterus depend on the position of the placenta and the umbilical cord. Third, the operation space for instrument within the uterus is limited and the quality of endoscope image falls off due to the cloudy amniotic fluid. Those conditions limit the design of fetal surgical instrument.

2.2 Fetus Stabilizing Mechanism and the Stabilizing Procedure

During the myelomeningocele surgical repair, the forces pushed or pulled to the fetus will change the position and the posture of the fetus. In order to prevent the fetus from

moving or rotating, the points supported to the abdomen side corresponding to the back of the fetus are required. Furthermore, it is necessary to support the breast of the fetus since the head is heavy. Considering of above two requirements, the designed fetal stabilizer should be enable supporting the abdomen side of the fetus from breast to the abdomen softly.

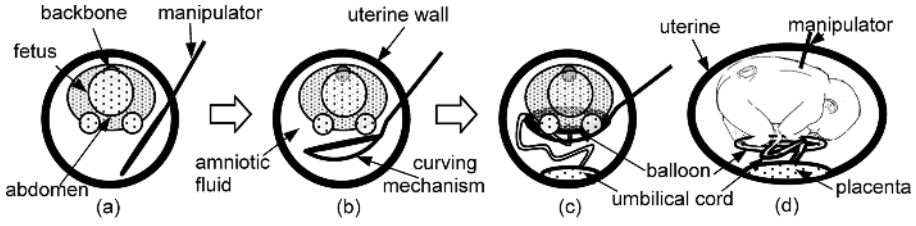


Fig. 1. Fetus supported by a balloon-based flexible manipulator. The insertion procedure shown in (a)-(c): (a) insert the manipulator straightly. (b) bend and adjust the curving mechanism to a optimal position. (c) pump the balloon by injecting physiological saline and stabilize the fetus. (d) viewed from left side.

To meet the requirements mentioned above, we developed a fetal stabilizing mechanism with a balloon that swells into a circle shape to support the fetus (Fig.1c,d). This balloon-based stabilizer is possible to support the fetus softly without any physical damage. We can fold the balloon and insert it into the uterus from a small opening hole. Furthermore, we use inject physiological saline from a syringe to swell the balloon. The injected amount and the pressure are controlled by PC. Even the balloon is broken during the operation it will not influence the fetus. The inserted part of the bending mechanism is covered by silicon so that the fetus and the placenta will not be injured.

The procedure of inserting the balloon-based stabilizer to the uterus and stabilizing the fetus with the ultrasound guidance is shown in Figure 1(a)-(c).

1) We insert the instrument with a shrunk balloon according to the position of the fetus. The spatial relationship of the fetus, the uterine wall and the instrument is observed using an ultrasonic diagnosis device.
2) The bending angle of the link part and the curving part of the manipulator are controlled by a PC according to position of the fetus. The bending and curving mechanisms are crooked and inserted into the required position with the guidance of intra-operative ultrasonic image.
3) When the silicon covered curving part arrive the abdomen of the fetus, we inject the physiological saline into the balloon and adjust the balloon to optimal position for stabilizing the fetus.

The circle-shape balloon supports the fetus near the hand and the foot. The umbilical cord passes by the balloon as shown in Fig.1c.

2.3 System Overview

The prototype of the fetal support manipulator is comprised of three units: the flexible manipulator unit with a balloon-based stabilizer and a curving mechanism; the PC for controlling the bending mechanism and the syringe; the ultrasonic device for viewing the intra-operative situation of the fetus and the instrument (Fig.2).

The balloon is connected with a syringe. The pressure is generated by an electric linear slider with a syringe (60 ml) similar to a syringe pump. The manipulator is held by the surgeon and controlled by the PC. The manipulator is performed under the guide of the ultrasonic image.

In general endoscopic surgery, since the viewing zone of endoscope is small and sometime the endoscope can not reach the narrow area, an ultrasonic image is used for assisting the endoscope surgery. The ultrasound is an ideal imaging modality for the diagnosis in obstetrics and gynecology. We use ultrasound to acquire the real-time intra-operative information. The position of the fetus and the umbilical cord could be identified when we insert the stabilizer into the uterus.

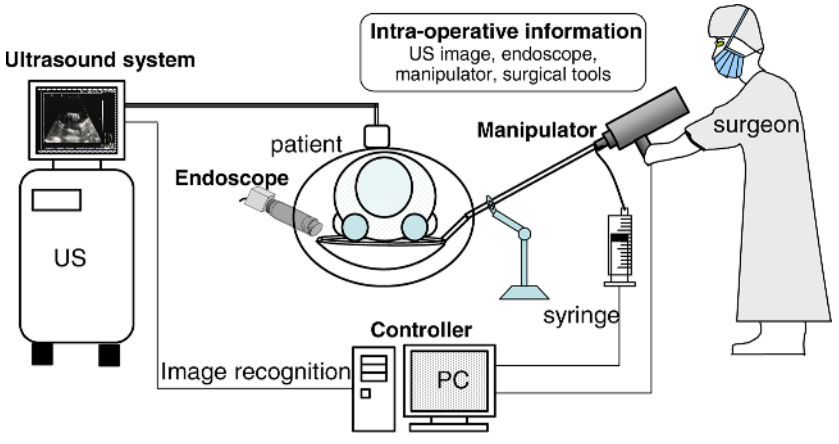


Fig. 2. Fetus support manipulator system for intrauterine surgery

2.4 Mechanism of the Manipulator

We put a tube inside the linkage for injecting the saline into the balloon (Fig.3c). The linkage mechanism made of stainless steel 6 mm in diameter. The length of the curving mechanism is determined by the size of balloon. Since the circumference of the fetus's body is about 15~20 cm and the diameter of the corresponding balloon is about 10 cm, we developed a curving mechanism with 15cm length which is suitable for enlarging the balloon.

In order to minimize the size of the insertion hole, we developed a curving mechanism that is comprised of a board spring and a wire-driven mechanism (Fig.3d). The balloon could be stored between the curving mechanism and the linkage part.

To the bending mechanism, we adopt a linkage-driven approach so that the manipulator enables a high stiffness (Fig.3b), durability and accurate performance than a wire-driven approach. Since the latex may trigger the fetus allergic reaction,

they will not be considered as material for the supporting device. The balloon is made of polyethylene so that fetus could be supported almost without pressure. The area in contact with the fetus becomes larger and the pressure to the fetus becomes smaller when the balloon is made bigger. However, the incision to the patient would be large. As shown in Fig. 3c, the diameter of the curving mechanism is 6 mm.

The tip part of the manipulator is sterilized by the autoclave (high-pressure steam), and connected with the actuator by draping mechanism.

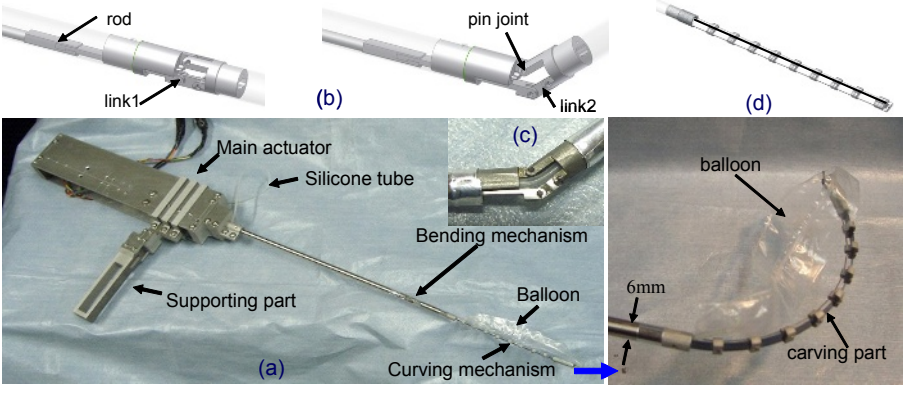


Fig. 3. Linkage bending mechanism, carving mechanism and balloon-based stabilizer

3 Experimental Results

The manipulator has two main functions: bending and curving the mechanism to the optimal position in the uterus; swelling the balloon to support the fetus. First, we evaluated the performance of the stabilizer curving and the linkage bending. Secondly, using the stabilizer to support the fetus, an experiment was carried our on a fetus model. We also evaluated the feasibility of the ultrasound guided manipulator.

3.1 Accuracy Evaluation of Stabilizer Curving Mechanism

We carried out experiment to evaluate the mechanical performance of the stabilizer bending mechanism. In order to swell the balloon to the required shape in fetus supporting, it is necessary to bend the curving mechanism and shorten the distance between the root and the tip part of the board spring. Furthermore, even the curving mechanism was not contacted the womb during the curving process, it is possible to touch the womb in recovering procedure if the hysteresis existed. The first set of tests aimed to measure the distance change from the root point to the tip end point of the curving linkage. We use a camera (QV-R51, CANON) to capture the pose of the curving process and measure the position of the tip end point. It was found that the hysteresis was existed and the error was 7 mm in maximum (Fig.4). This result showed the accuracy of curving mechanism is sufficient in stabilizing the fetus.

3.2 Accuracy Evaluation of Linkage Bending Performance

The second set of tests assessed the performance of the linkage bending mechanism. We inserted and adjusted the stabilizer corresponding to the position of the placenta using the bending mechanism. It is necessary to ensure the bending range of the linkage mechanism. We evaluated the accuracy of the bending angle from -45° to 45° and compared the results with the theoretical values. It was found that the measured values were satisfied with the theoretical value (Fig.5). The maximum error of the bending angle was 5° and the standard deviation was 1.6°.

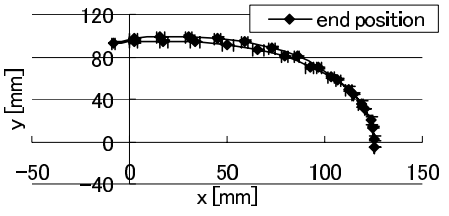


Fig. 4. Tip end point positions of the manipulator

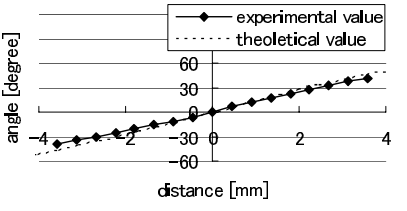


Fig. 5. Relationship between the translation distance and the bending angle

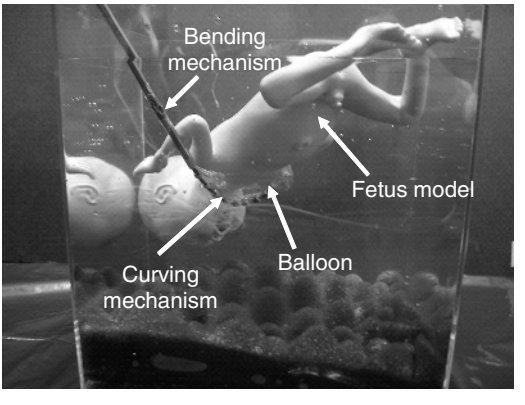


Fig. 6. Manipulator stabilizes the fetus model softly

3.3 Fetus Supporting Experiment Using Fetus Model

The flexible fetal supporting manipulator was applied to a fetus model stabilizing experiment. The fetus model weighs 600g with volume of about 600 cm^3 and length of 30cm. The long radius of the abdomen is about 6 cm. The fetus model was suspended in a water tank. We inserted the manipulator from the side of the fetus, bending the linkage mechanism and crooking the stabilizing mechanism. The balloon was swelled and adjusted to support the fetus from the abdomen side. Figure 6 showed that the developed manipulator could support the fetus model and stabilize the posture. The picture also showed that only the balloon contacted with the fetus and the metal curving mechanism was not in contact with the fetus. The balloon could

stabilize the fetus and prevent it from moving or rotating. The experiment showed that the fetus model could be supported and lifted up sufficiently using the manipulator.

3.4 Ultrasound Guided Manipulator Insertion

We carried out experiments to evaluate the feasibility of the ultrasonic image guidance to the insertion of the manipulator. The situation of the fetus model in the action of the manipulator is also identified using the ultrasonic device.

The first set of tests is to evaluate the ultrasonic image of the manipulator without placing the fetus model. Figure 7a showed the situation when the balloon was shrunk. The curving mechanism was observed clearly. Figure 7b showed that both of the mechanism and the balloon were observed clearly after the balloon was swollen up.

The second set of tests is to use fetus model to similar an implementation of fetal surgery with the developed manipulator. Figure 7c illustrated that the fetus model and the manipulator could be observed. We adjusted the stabilizer to the optimal position to support the fetus (Fig.7d). The manipulator and the curving mechanism could be observed under the fetus model clearly.

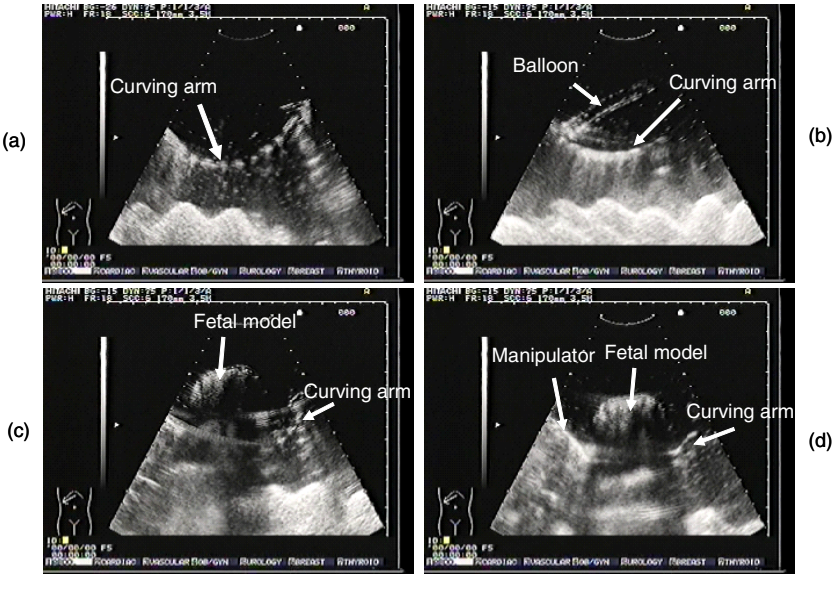


Fig. 7. Ultrasound guided manipulator insertion in stabilizing the fetus model (a) the curving mechanism was observed clearly. (b) both of the curving mechanism and the swollen balloon were observed. (c) the fetus model and the manipulator is observed . (d) The manipulator and the curving mechanism could be observed under the fetus model.

4 Discussions and Conclusion

The prototype manipulator has the potential to be successfully inserted into the uterus during the endoscope intrauterine surgery, while the balloon-based stabilizer has the

potential to support the fetus softly. The experimental results demonstrated feasibility of the developed manipulator. First, the hysteresis of the curving mechanism showed as small as a 7-mm error and the accuracy of the linkage bending mechanism revealed the standard deviation of bending angle was just 1.6°. Second, the fetus model could be supported and lifted up by the balloon based stabilizer. Third, the manipulator could be controlled and the curving mechanism with balloon in the stabilizer could be clearly observed under ultrasound guidance.

In the minimally invasive fetoscopic surgery, the balloon is expected to be larger for fetus stabilization unlike its insertion sites which should be made as small as possible. Then, it is necessary to improve the method of balloon folding and storing.

The use of ultrasound was not enough to give information below the model (Fig. 7c), although it clearly visualized the manipulator and the fetus model as well. To achieve a full intra-operative guidance, the endoscope could be also available.

In conclusion, we reported a newly developed method of supporting the fetus during the endoscope intrauterine surgery. The flexible manipulator could be controlled to reach the optimal position in the uterus and the balloon-based device could stabilize the fetus softly. The manipulator is potentially helpful for minimally invasive intrauterine surgery, although further improvements and experiments are required.

Acknowledgment

This study was supported in part by the Grant-in-Aid of the Ministry of Education, Culture, Sport, Science and Technology in Japan, and Grant-in-Aid of the Ministry of Health, Labour and Welfare in Japan.

References

[1] D.C. McCullough, D.L. Johnson, “Myelomeningocele repair: technical considerations and complications,” Concepts Pediat Neurosurg, Vol.8, pp29-40, 1988.

[2] D.S. Heffez, J. Aryanpur, G.M. Hutchins, J. Freeman, “The paralysis associated with myslomeningocele: Clinical and experimental data implicating a preventable spinal cord injury,” Neurosurgery Vol. 26, pp.987-992, 1990.

[3] O.S. Aaronson, et al. “Robot-assisted endoscopic intrauterine myelomeningocele repair: A feasibility study,” Pediatric Neurosurgery, Vol.36 No.2, pp.85-89, 2002.

[4] K.C. Oberg, A.E. Robles, et al. “ Endocsopic intrauterine surgery in primates,” Surgical Endoscopy, Vol.13, pp420-426, 1999.

[5] M. Dinh, J. Bruner, “Fetal Stabilizer for Intrauterine Surgery”, Report of Vanderbilt University.

[6] Tsubouchi et al. “Development of the drawn-in type stabilizer for a fetus operation,” The 14th conference of Japan computer-aided-surgery society, pp.45-46, 2005.

Recovery of Surgical Workflow Without Explicit Models

Seyed-Ahmad Ahmadi[1], Tobias Sielhorst[1], Ralf Stauder[1], Martin Horn[1], Hubertus Feussner[2], and Nassir Navab[1]

[1] Chair for Computer Aided Medical Procedures (CAMP), TU Munich, Germany
[2] Chirurgische Klinik und Poliklinik, Klinikum Rechts der Isar, TU Munich, Germany

Abstract. Workflow recovery is crucial for designing context-sensitive service systems in future operating rooms. Abstract knowledge about actions which are being performed is particularly valuable in the OR. This knowledge can be used for many applications such as optimizing the workflow, recovering average workflows for guiding and evaluating training surgeons, automatic report generation and ultimately for monitoring in a context aware operating room.

This paper describes a novel way for automatic recovery of the surgical workflow. Our algorithms perform this task without an implicit or explicit model of the surgery. This is achieved by the synchronization of multidimensional state vectors of signals recorded in different operations of the same type. We use an enhanced version of the dynamic time warp algorithm to calculate the temporal registration. The algorithms have been tested on 17 signals of six different surgeries of the same type. The results on this dataset are very promising because the algorithms register the steps in the surgery correctly up to seconds, which is our sampling rate. Our software visualizes the temporal registration by displaying the videos of different surgeries of the same type with varying duration precisely synchronized to each other. The synchronized videos of one surgery are either slowed down or speeded up in order to show the same steps as the ones presented in the videos of the other surgery.

1 Introduction and Related Work

Surgical workflow recovery is based on the hypothesis that surgeries of one certain type can be compared to each other. In general, these procedures are never performed in exactly the same way, since the experience of the surgeon and the anatomy of the patient vary and complications may occur. Despite these differences, surgeries of one type generally share a common sequence of events, which is typical for the procedure. As Herfarth denotes, a thorough and smooth integration of surgical workflow in the clinical framework is becoming more and more necessary. He postulates that surgical units have to become specialized and dedicated to only one type of surgery up to a specialization on individual organs [1]. The recovery of workflow is one of the crucial tasks towards the long-term goal of opinionated consultancy inside the OR, which has been drafted by experts during discussions in the OR2020 workshop of 2004 [2]. The benefits of a

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 420–428, 2006.

system which is aware of and reactive towards the current workflow are for the medical as well as for the administrative side of a hospital.

Riedl et al. postulate in their work more transparency of events within the OR in order to improve efficiency of room and tool usage [3]. Aggarwal et al. [4] have discussed the possibilities of surgery simulation for training purposes, since it has the potential to reduce the occurrence of adverse events. Their OR simulation environment records various data streams, both video and audio, for the assessment of surgical and team skills through expert surgeons in debriefing sessions. Rosen et al. [5] read data from force/torque sensors at laparoscopic instruments and analyse video data about the tool/tissue interaction for the segmentation of movements using hidden markov models (HMMs). Their objective is to evaluate surgical skills. Lin et al. [6] follow a similar approach and propose a system for automated workflow recovery based on motion segmentation with linear discriminant analysis, bayes classifiers and HMMs. Their framework is the analysis of robot-assisted movements with the Intuitive Surgical's da Vinci system during a suturing task. Their aims are the evaluation of surgical skill levels and training of inexperienced surgeons. Strauss et al. [7] have created manual protocols of endoscope usage time, pose changes and instrument changes in order to prove the usability and necessity of an ontology for the description of surgical workflow. They postulate and conceptualize a mechatronic assistance system for endoscope guidance. Sielhorst et al. [8] synchronize 3D hand movements of a physician using Dynamic Time Warping (DTW). They use this in an AR setting for the synchronized playback of movements from trainees and experts for training and evaluation purposes.

While [5] and [6] have been able to segment single and isolated tasks within a whole surgical procedure, we see many benefits of a system that is able to understand the overall workflow of a complete surgery. In this paper, we propose a DTW-based system, which is specialized, but not confined to the classification and labelling of workflow steps within a complete minimally-invasive procedure. Approaches using Hidden Markov Models allow for modelling the workflow as a non-linear procedure, but they need the definition of a limited number of well-defined states. We present a complementary approach that allows for discovery of these states from a large number of signals acquired during surgeries of the same type. This paper describes the implementation of our approach and first results on six minimally invasive surgeries.

2 Surgical Environment

As an exemplary surgery for the workflow retrieval we have chosen cholecystectomy. 95% of the removal of the gall bladder are performed laparoscopically with a very low conversion rate. In spite of its complexity, the frequent performance of the surgery and its stable outcome make it a practical type of surgery for our kind of analysis. In order to relate to the materials and methods that have been used for workflow recovery in the course of this paper, a short verbal description of the workflow follows.

After anaesthesia and further preparation of the patient, like disinfection of the abdominal region, a small incision is made at the umbilicus or navel of the patient. The abdomen of the patient is inflated with CO2 in order to create an abdominal cavity. This marks the beginning of the procedure to us. When the inflation is sufficient, the trocars and the laparoscope for minimally invasive surgery are inserted into the abdomen. A metal rod is inserted through one of the trocars, the liver is pulled up and the gallbladder attached beneath is revealed.

Using a grasper and a dissecting device, the surgeon then identifies the positions of the gallbladder's two main vessels, the common bile duct and the bile artery. These vessels are uncovered from adhesions and adjacent tissue. When sufficiently dissected, they are clamped using a clipping device and subsequently cut using a pair of laparoscopic scissors. As a next step, the gallbladder is removed from the liver. In order to separate the gallbladder from the liver, high-frequency current is applied in two modes, cutting and coagulation mode. The metal tips of most instruments serve as electrodes, which are able to dissect the tissue electrically. During the removal of the gallbladder, it is likely that bleedings are caused in the so-called liver bed area, which is a large portion at the lower side of the liver where the gallbladder used to reside. After the gallbladder has been removed completely, these bleedings are stopped using high-frequency current in coagulation mode. In order to retrieve the gallbladder, a retraction sac is inserted and the gallbladder is put into the plastic bag with two or three graspers. One of the trocars is pulled out and the plastic bag is retracted through the umbilical port. After removal, the trocar is inserted again and in a last and final control phase, the situs is once more controlled for bleedings. The remaining vessel stumps are controlled for correct and clean clipping. A drainage tube is inserted in order to lead away the saline irrigation fluid after surgery. All instruments and trocars are retrieved and the four incisions are closed with stitches. The beginning of the abdominal suturing marks the end of laparoscopic activity and thus the end of the workflow that we are considering for this work.

This sequence of events is specific to a standard laparoscopic cholecystectomy. Although variations in the duration of certain phases may occur due to different complications, the general workflow is systematic and thus reproducible.

3 Algorithms and Methods

3.1 Principle of Workflow Recovery

In minimally-invasive surgeries, the laparoscopic instruments used by the surgeon strongly correlate with the underlying workflow. We account for the ongoing actions during the procedure with a binary model for the instrument usage:

$$u(t) = \begin{cases} 1 & \text{if the instrument is used at time } t, \\ 0 & \text{if the instrument is not used at time } t. \end{cases}$$

A series of multi-dimensional state vectors over time takes into respect that several instruments can be used simultaneously. Currently, we observe the usage of 17 different laparoscopic instruments.

Our novel way of workflow recovery is based on the following four steps:

1. **Synchronization of events.** Using the DTW algorithm, we synchronize different surgeries with variable duration in a non-linear manner. Due to the synchronization of signals, the underlying events and workflow steps are synchronized automatically as well.
2. **Creation of an average surgery.** An average surgery is created. It reveals events that are common to all procedures.
3. **Identification of workflow phases.** Common events within the average surgery trigger the start and end points of workflow phases.
4. **Obtaining workflow phases for the original surgeries.** Using the time mapping from each surgery onto the average, workflow phases for each original surgery can be identified at the end of the procedure.

3.2 Classical DTW

The DTW algorithm is able to perform the non-linear synchronization that is necessary for the mapping of our surgery data. DTW was most probably originated in the field of speech recognition [9] and was further used in the synchronization of chemical batch processes [10] as well as in many other fields such as motion detection and segmentation [11] [12] or synchronization [8].

In this paper, only the basic principles of the DTW and the extensions which were necessary for this project shall be made clear. For details and further explanations of the algorithm, it is recommended to refer to [9], [10] and [8].

The DTW is a recursive function, which aims at providing correspondencies reducing the distance between two trajectories A and B. For each time point i of curve A, the distance is calculated to each time point j of curve B and denoted in a 2D DTW matrix field.

The recursive definition of the DTW algorithm [9] and the principles of dynamic programming [13] have the effect that a trench of lowest values occurs while the DTW matrix is filled. Backtracking this trench yields the so-called warp path $h(k)$, where k denotes a newly established, common timeline of the two curves. This warp path is the optimal linkage between the two timelines i and j, so that the curves A and B are mapped onto each other according to their similarities as best as possible.

By using a multi-dimensional distance function $d(i, j)$, we are able to compare the state vectors of our surgeries and obtain the desired optimal DTW time mapping (cf. Figure 1). For our experiments, we made use of the standard Euclidian distance.

3.3 Creating an Average Surgery

The creation of an average surgery is crucial for our proposed method of workflow recovery. In order to create such an average, more than two surgeries need to be synchronized onto a common timeline. Although the classic DTW algorithm only allows synchronizing two curves at a time, Wang and Gasser have overcome this

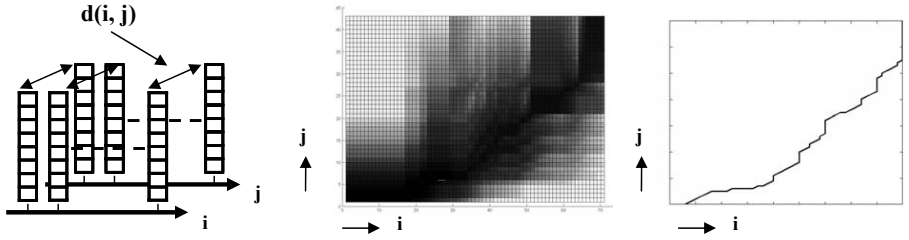


Fig. 1. DTW warp paths synchronize the time axes i and j of two surgeries so that their state vectors are optimally aligned.

problem by making use of the transitive property of the DTW [14] [15]. Using their method, we synchronize an arbitrary number of surgeries by applying the DTW several times according to the following scheme:

- Out of a set of m surgeries OP_i, $(i = 1 \dots m)$, one surgery is chosen as a reference curve OP_{Ref} and all other surgeries are mapped onto the reference using the multi-dimensional DTW. Thus, we obtain $(m-1)$ warp paths $h_i(t)$. $h_i(t) = DTW(OP_i, OP_{Ref})$
- The warp paths are interpolated and averaged, yielding a warp path h(t) between the reference curve and a common, average timing. $h(t) = \frac{1}{m} \sum_{i=1}^{m} h_i(t)$
- In the third step, we calculate the desired warp paths $u_i(t)$ between the average timing and the timing of each original surgery. These so-called shift functions are calculated by taking the inverse of $h(t)$ as the argument of the warp paths $h_i(t)$: $u_i(t) \equiv h_i\left(h^{-1}(t)\right)$
- Using these shift functions, the instrument state vectors of each surgery can be arranged onto the new, average timeline. Subsequently, the instrument vectors themselves are averaged for each timestamp, yielding the so-called structural average OP_{Avg}: $OP_{Avg}(t) = \frac{1}{m} \sum_{i=1}^{m} (u_i(t))$

For the reference curve, we chose the one with a duration that is closest to the average duration of the five surgeries. Due to the averaging process, state vectors in OP_{Avg} feature non-binary values as opposed to our original model. These values can be interpreted as quasi-probabilities. Let us assume that at t_a, n out of the m surgeries from the averaging set had a certain instrument in use. Then, the value for this instrument in the average will be: $OP_{Avg}(t_a) = \frac{n}{m}$

These values are not a probability in the axiomatic sense, but they give an idea of how well an instrument could be synchronized by the algorithm. Furthermore, they reflect how well the surgeries correspond to each other at t_a. More importantly, these values reveal events, which are typical to the surgical procedure.

3.4 Weighting of Signals From the Curves

It makes sense to bias the DTW synchronization towards instruments which have a high relevance for the underlying workflow. For example, the clipping device

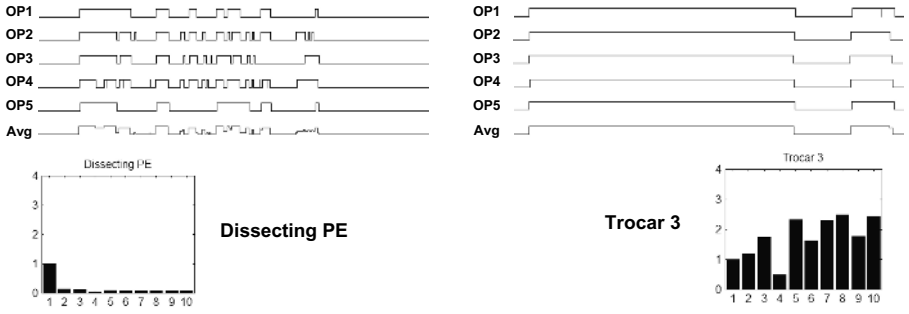


Fig. 2. Weighting of instruments according to their workflow relevance: the better the synchronization, the higher the weight.

clearly indicates the work step of clamping a vessel, while the dissecting device is used as an all-round instrument in many different phases of the surgery.

Mathematically, the bias is achieved by assigning different weights to the instruments during the calculation of the distance function. The problem of weighting certain variables over others in a multi-dimensional setting has been encountered before by the group around Kassidas et al. [10].

Adopting their method, we perform iterative calculations of an average surgery. After every round, each instrument from the state vector is compared to its synchronized equivalent. A new weight is then assigned and subsequently normalized.

$$W(j,j)=\left[\sum_{i=1}^{n}\sum_{k=0}^{t_{avg}}\left[OP_i(k,j)-OP_{Avg}(k,j)\right]^2\right]^{-1},\ W\left\{n_{Instr}/\left[\sum_{j=1}^{n_{Instr}}W(j,j)\right]\right\}$$

In simple words, the rule is to give higher weight to those instruments, which could be synchronized very well and lower weight to instruments with only poor synchronization (cf. Figure 2).

4 Experiments and Results

4.1 Workflow Retrieval

For the evaluation of our developed algorithms, six surgeries were recorded by courtesy of our clinical partner. The average surgery was calculated with five surgeries from the set and after ten iterations of weight updating. The sampling frequency for the instrument state vectors was 1 Hz. The state vectors were manually read out from the laparoscopic video. The sixth surgery was then synchronized to the average in order to evaluate the quality of workflow retrieval for surgeries, which are unknown to the system. The resulting average surgery OP_{Avg} can be seen in Figure 3.

The curve segment from OP_{Avg} marked with a dashed ellipse demonstrates that common events are carved out while uncommon events tend to be attenuated but not neglected within the average surgery. This happens only due to

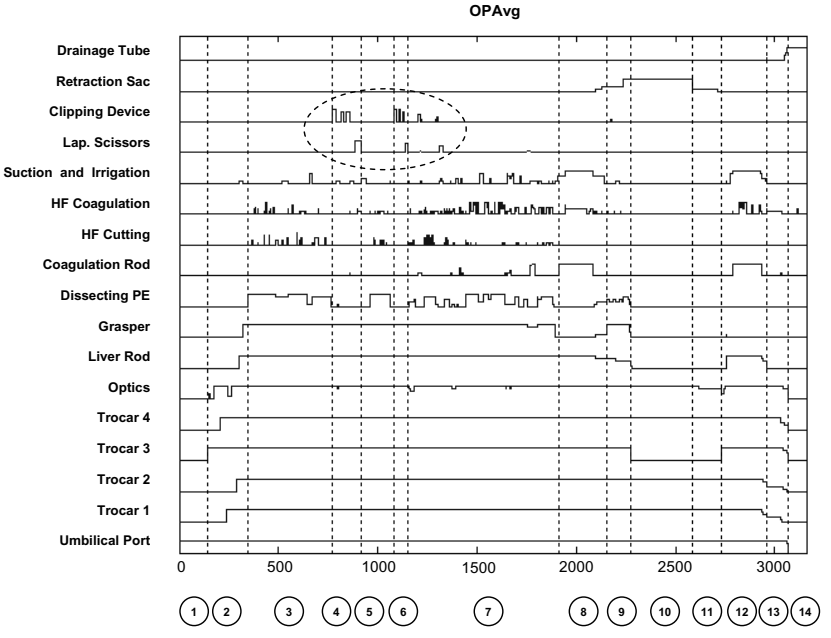


Fig. 3. Average surgery with 14 workflow phases

the synchronization and without prior knowledge or models on the workflow. A short explanation shall substantiate this assertion. For every patient, two main gallbladder vessels have to be severed. This is represented by three high peaks for the clipping device and one high peak for the scissors. For some patients, these vessels have branched, which requires additional clipping and cutting. Since these additional events are rather uncommon, the following spikes have lower amplitudes or lower quasi-probabilities.

State changes from zero to one or one to zero within OP_{Avg} were defined as trigger events for workflow phases. Using these triggers, 14 workflow phases could be identified within the average curve. The segmentation of these phases in the multi-dimensional curves can be seen in Figure 3. The same trigger events were identified within each original surgery. Thus, two trigger time points exist for each phase in an original surgery. The first is the original time stamp t of the phase beginning. The second one, t^*, is obtained through the average surgery OP_{Avg} and the shift function $u_i(t)$, which connects the timeline of OP_i with the one of OP_{Avg}.

A deviation of up to five seconds between these timestamps was tolerated. Deviations higher than five seconds were classified as a wrong phase detection. In six procedures with 13 trigger events each, the proposed system was able to identify 92% of the events correctly. Moreover, 83% of the correctly classified phases were detected with a precision of one second or less. These results demonstrate that our approach is reliable and promising. The video presentation shows the user interface and some exemplary results.

5 Conclusion

Workflow recovery is a key competency of context-sensitive systems and workflow-aware operating rooms. We have shown that our approach of workflow recovery is able to identify common phases in different surgeries of the same type even under the conditions of a real surgery. For this task, an average surgery is computed from a set of signals from exemplary surgeries, which already provides the desired key information for many workflow specific applications. The average curve is achieved by synchronizing multiple procedures and it reveals key events of the procedure. The change of workflow phases can be identified with a reliability of 92% with a tolerance of 5 seconds. Each of the signals including the video streams from the laparoscope and the video stream of our three external cameras can be shown simultaneously for two or more surgeries. The fine-grained synchronization of the algorithm allows for stretching the playback speed in a way that the surgeries are visualized according the workflow phases. Therefore, each single frame of all 24 videos is appropriately labelled with synchronized workflow information for further studies. The playback is useful for thorough and unprecedented analysis of surgical workflow, educational and training purposes and evaluation of surgical skills.

References

1. Herfarth, C.: 'lean' surgery through changes in surgical workflow. British Journal of Surgery **90** (2003) 513–514
2. Cleary, K., Chung, H.Y., Mun, S.K.: Or 2020: The operating room of the future. Laparoendoscopic and Advanced Surgical Techniques **15** (2005) 495–500
3. Riedl, S.: Modern operating room management in the workflow of surgery. spectrum of tasks and challenges of the future. Der Chirurg **73** (2002) 105–110
4. Aggarwal, R., Undre, S., Moorthy, K., Vincent, C., Darzi, A.: The simulated operating theatre: comprehensive training for surgical teams. Qual Saf Health Care **13** (2004) 27–32
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. Comput Aided Surg. **7** (2002) 49–61
6. Lin, H.C., Shafran, I., Murphy, T.E., Okamura, A.M., Yuh, D.D., Hager, G.D.: Automatic detection and segmentation of robot-assisted surgical motions. In: MICCAI 2005. Volume 3749., Springer (2005) 802–810
7. Strauss, G., Fischer, M., et al., J.M.: Workflow analysis to assess the efficiency of intraoperative technology using the example of functional endoscopic sinus surgery. HNO (2005)
8. Sielhorst, T., Blum, T., Navab, N.: Synchronizing 3d movements for quantitative comparison and simultaneous visualization of actions. In: Proc. IEEE and ACM International on Mixed and Augmented Reality (ISMAR). (2005)
9. Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans. Acoust. Speech Signal Process. **26** (1978) 43–49
10. Kassidas, A., MacGregor, J.F., Taylor, P.A.: Synchronization of batch trajectories using dynamic time warping. AIChE Journal **44** (1998) 864–875
11. Li, H., Greenspan, M.: Multi-scale gesture recognition from time-varying contours. Tenth IEEE International Conference on Computer Vision (ICCV) **1** (2005) 236–243

12. Bobick, A.F., Wilson, A.D.: A state-based approach to the representation and recognition of gesture. IEEE Transactions on Pattern Analysis and Machine Intelligence **19** (2005) 1325–1337
13. Bellman, R.E., Dreyfus, S.E.: Applied Dynamic Programming. Princeton University Press (1962)
14. Wang, K., Gasser, T.: Alignment of curves by dynamic time warping. Annals of Statistics **25** (1997) 1251–1276
15. Gasser, T., Wang, K.: Synchronizing sample curves nonparametrically. Annals of Statistics **27** (1999) 439–460

Comparison of Control Modes of a Hand-Held Robot for Laparoscopic Surgery*

Oliver Tonet[1,2,**], Francesco Focacci[2,1], Marco Piccigallo[2,3], Filippo Cavallo[1,2], Miyuki Uematsu[4], Giuseppe Megali[1,2], and Paolo Dario[1,2]

[1] CRIM Lab, Scuola Superiore Sant'Anna, Pisa, Italy
oly@sssup.it
[2] EndoCAS Center for Computer-Assisted Surgery, University of Pisa, Italy
[3] IMT Institute for Advanced Studies, Lucca, Italy
[4] Major in Bioscience and Biomedical Engineering, Waseda University, Tokyo, Japan

Abstract. Teleoperated robots for minimally invasive surgery make surgeons loose direct contact with the patient. We are developing a hand-held, dexterous surgical robot that can be controlled with one hand only, while standing at the operating table. The instrument is composed of a master part (the handle) and a slave part (the tip). This work compares the performance of different control modes, i.e. different ways to map the degrees of freedom of the handle to those of the tip. We ask users to drive the tip along complex trajectories in a virtual environment, using the real master to drive a simulated slave, and assess their performance. Results show that, concerning time, users with no training in laparoscopy prefer a direct mapping of position and orientation, like in free hand motion. However, users trained in laparoscopy perform equally fast with our hand-held robot and, concerning precision, make a smaller number of errors.

1 Introduction

Minimally invasive surgery procedures bring numerous benefits to the patient, but severely hamper surgeons' perception and motor skills. Technology can provide methods and devices aimed at restoring – and possibly augmenting – visual, haptic and tactile feedback, motor coordination and dexterity: stereo endoscopes and displays, teleoperated master/slave systems, robotic interfaces and surgical instruments with many degrees of freedom (DoFs). Teleoperated systems, like the da Vinci Surgical System [1] by Intuitive Surgical Inc., bring some advantages at the cost of longer setup times and of cluttering the already crowded operating table, pushing the surgeon away from the patient: the surgeon operates at a console and only the robot is in direct contact with the patient. Experienced

* This work has been supported in part by the FIRB-2001 Project *ApprEndo* (no. RBNE013TYM) and by *EndoCAS*, the Center of Excellence for Computer-Assisted Surgery (COFINLAB-2001 no. CLAB01PALK), both funded by MIUR, the Italian Ministry of Education, University and Research.

** Correspondence author.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 429–436, 2006.

surgeons tend to agree that in many procedures the benefits provided by teleoperated systems are not really needed during the whole surgical procedure, and they tend to prefer the traditional hands-on approach for routine tasks.

Led by these considerations, we are developing a hand-held surgical robot that can be operated by a surgeon with one hand only, while the other hand is free to use a traditional surgical instrument. The robot consists of a master part (the handle) and a 6-DoF slave part (the tip), connected together [2]. Selecting the best way to control the tip by means of the handle is one of the major issues in the design. A *control mode* is defined as a way to map the DoFs of the handle to the DoFs of the tip [3]. In this paper we present an experiment aimed at selecting the control mode that provides the most intuitive and efficient way to steer the instrument. We have developed a teleoperated system composed of custom handles mounted on a haptic interface that also serves as 6-DoF digitizer. On this system, we implemented an exercise mimicking a surgical gesture requiring high dexterity, namely knot tying, and asked subjects to perform the exercise with four different control modes in order to select the most efficient. Two control modes are feasible on a hand-held instrument, while two require kinematic separation between master and slave. The latter have been included to compare the performance of the hand-held configuration to master/slave teleoperation systems.

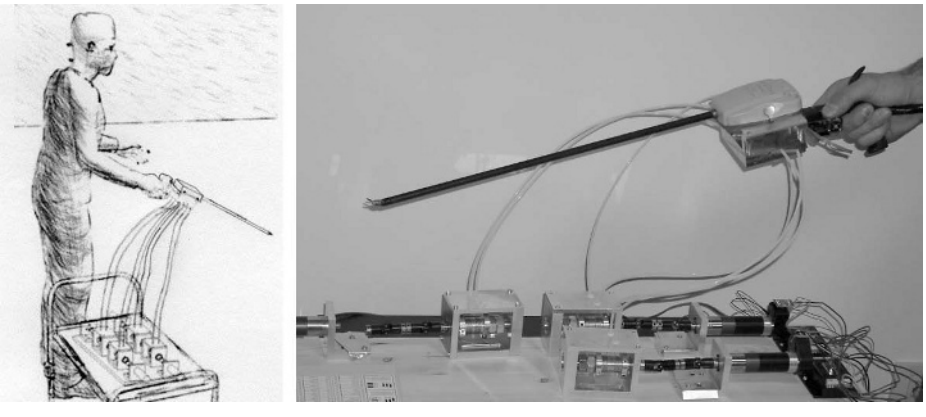


Fig. 1. Concept drawing of the lightweight hand-held laparoscopic robot (left) and the first prototype (right), using the EndoWrist of the da Vinci Surgical System as end-effector

2 Methods

2.1 Control Modes

For the comparative evaluation, we selected four control modes, obtained by switching two control parameters independently from each other. The first parameter is defined *PCM* (position control mode) and its values are *Tip* or *Handle*; the second parameter is called *ACM* (angle control mode) and its values are

Absolute or *Relative*. PCM denotes how the position of the master is mapped to the position of the slave, while ACM specifies how the angular DoFs at the instrument tip are controlled. PCM=Tip means that the position of the master is directly mapped to the position of the tip. In other words, moving the master in one direction causes the tip to move in the same direction. By setting PCM=Handle we introduce a fulcrum constraint at the trocar position, hence moving the master in one direction causes the tip to move in the opposite direction (as in laparoscopic surgery). By setting ACM=Absolute, the orientation of the tool tip in 3-D space will be exactly the same as the orientation of the handle. By setting ACM=Relative, the angles at the tip joints are controlled incrementally through a joystick (moving the joystick in one direction makes the corresponding tip angle increase in the same direction). By combining the two PCMs with the two ACMs we obtain four different control modes that can be evaluated on our platform: *H_A* that means (PCM=Handle, ACM=Absolute), *T_A* (PCM=Tip, ACM=Absolute), *H_R* (PCM=Handle, ACM=Relative), and *T_R* (PCM=Tip, ACM=Relative). E. g. teleoperated master/slave systems like the da Vinci Surgical System have *T_A* control mode, whereas our hand-held robot has *H_R*. Of course, a hand-held instrument will have PCM=Handle, while PCM=Tip requires separate kinematics for the master and slave parts.

2.2 Teleoperated System with Simulated Robot

Figure 2 shows the architecture of the system used for the experiments. For the master system, we fabricated two types of handle, one for each ACM. To control the slave angles of the instrument in the ACM=Relative configurations, a joystick, driven by the user's thumb, has been mounted on the foil grip used as ergonomic handle. The joystick is a two-axis potentiometer and communicates with the host PC through an electronic board connected to the serial port. A switch has also been mounted in the foil grip in order to implement the stem roll movement in the virtual environment. For the ACM=Absolute configuration we fabricated a tweezers-like handle. The handles were connected to a Phantom Premium 1.0 by Sensable Inc., which is a 6 DOF localizer and 3 DOF force feedback interface. We used the Phantom to measure the master position; we only used the orientation encoders when ACM=Absolute.

To simulate PCM=Tip the handle is free to move in space and the position of the Phantom end-effector is mapped to the instrument tip position directly. To replicate PCM=Handle, the handle (foil grip or tweezers) is attached to the Phantom end-effector and the Phantom stylus, extended with an aluminum cylinder, is inserted through a trocar on an abdomen replica. Since the length of the instrument and the position of the trocar are known, we can compute the tip position by measuring the position of the Phantom stylus. The four configurations are depicted in Fig. 3. It is worth noting that in *H_A*, an additional gimbal ring was needed, allowing to control the handle orientation independently from the aluminum cylinder, constrained by the trocar.

The slave robot kinematics has been simulated in a virtual reality environment (see Fig. 4.right). We derived our robot design from the end-effector of

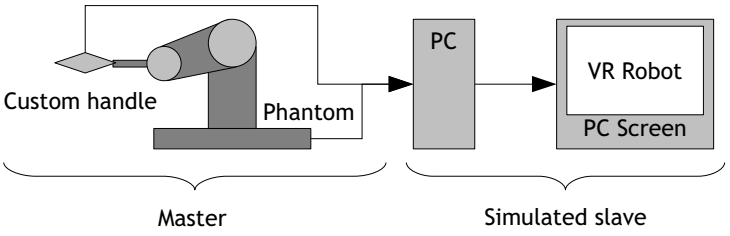


Fig. 2. The architecture of the system, with sensorized master and simulated slave

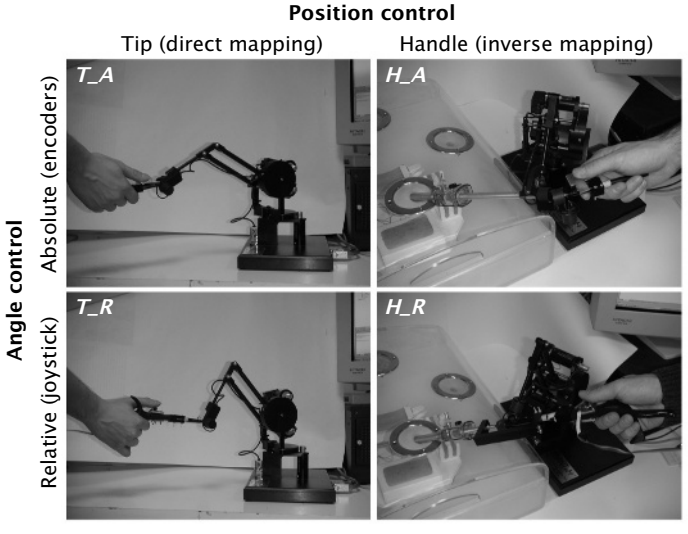


Fig. 3. The four control modes and the experimental setup

the da Vinci, called EndoWrist. We exported 3D objects from the CAD models of the mechanical components and assembled them in a 3D scene graph using the OpenGL Optimizer library. We calculated the inverse kinematics of the EndoWrist in order to map the end-effector position and orientation correctly and reproduce realistic animation in the virtual environment.

2.3 Performance Assessment

To compare the four control modes we asked test subjects to reproduce, by means of the different interfaces, a gesture that follows a complex path in 3-D space, thus requiring high dexterity and use of all 6 DoFs of the instrument tip. The trajectory to reproduce was created by sampling, at 60 Hz, the coordinates of the instrument tip while a surgeon was mimicking a knot-tying gesture on the platform itself. The data were smoothed with a low-pass filter, interpolated by means of a cubic spline, and downsampled to obtain a final trajectory composed of 100 points.

In the virtual environment the trajectory is shown as composed of a sequence of spheres that must be pierced by means of a needle placed in the slave forceps. A ring, perpendicular to the needle, is displayed as a visual aid to identify the needle orientation, as depicted in Fig. 4. Only a few of the spheres composing the trajectories are shown at the same time, to keep the display clear. One sphere is marked in bright yellow: this is the active sphere. To cross the active sphere successfully with the ring, the user must slide the needle past the sphere center, without touching the sphere with the ring, and keeping the ring normal to the direction tangent to the trajectory. Actually, a tolerance angle of 60 deg has been used. When the active sphere has been crossed, the following sphere is marked as active and so on, until the whole trajectory has been correctly followed. As additional requirement, to cross the actual active sphere successfully, the two previous spheres must also have been crossed successfully. These spheres are highlighted when they have been crossed. Moving away from the trajectory resets the task and requires the user to cross the two previous spheres again. In the following, we will refer to this latter condition as *off* state; conversely, we define the *on* state when the user is crossing the spheres correctly. The position and orientation of the instrument tip, together with a time stamp and all interactions with the spheres are logged on file for offline analysis.

A group of five subjects with only some experience on a laparoscopic trainer was asked to complete two trials with each of the four control modes. The trials were performed during two days, with no predefined order, allowing the subject at least 10 minutes of rest between the trials. We decided not to involve expert surgeons at this stage in order to avoid biasing the results with dexterous movement schemes expert surgeons have acquired during prolonged practice with laparoscopic instruments.

2.4 Data Analysis

The period of the simulation cycle is between 15 and 16 ms. The position and orientation of the tool was sampled from the Phantom interface at every cycle, sufficient to describe human gesture [4], especially fine movements. performed during surgery [5]. The acquired data was interpolated and resampled at a constant rate of 70 Hz. Only low-frequency components are present in surgeon's movements. However due to artifacts and some error of instruments, 3D position data is also contaminated by high-frequency noise. Since our data analysis involves first, second and third derivatives, the acquired data are off-line filtered using a numerical fourth-order low-pass Butterworth filter, with cut-off frequency of 15 Hz.

For the evaluation of the exercises, we have computed the following parameters: total time spent (t), total time spent in *on* state (t_{on}), total time spent in *off* state (t_{off}), percentage of time spent in off state ($t := t_{\text{off}}/(t_{\text{on}} + t_{\text{off}})$), number of errors, mean velocity, mean acceleration, and normalized jerk [6].

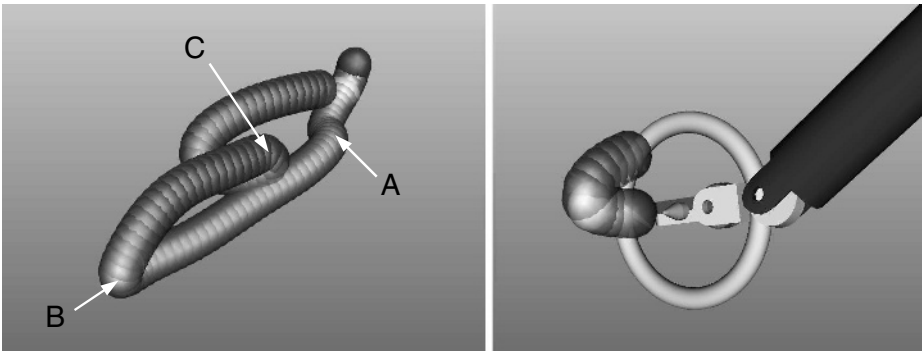


Fig. 4. The trajectory mimicking a knot-tying task as displayed in the virtual environment (left) and the simulated slave robot approaching the active sphere (right)

3 Results

Figure 5 shows the progression of the users during the exercises, separately for each control mode. The five users are called U1-U5. We have chosen to plot only the second trial in order not to overload the figure; the first trial has roughly similar shape. The horizontal axis shows which of the 100 spheres that build the trajectory the user was interacting with. The vertical axis shows the time, in s. The nearly horizontal segments correspond to parts of the trajectory where the user was progressing, crossing the spheres rapidly. The sloped segments correspond to difficult points in the trajectory, where the users spent much time in advancing to the next sphere. Comparing performance of users we can see that there are mainly three difficult points, roughly corresponding to spheres number 15, 45, and 70. These points correspond to passages with higher curvature, as highlighted in Fig. 4.left with the letters A, B, and C. We can also see that passage B is more difficult with ACM=Absolute (encoders), and passage C is more difficult with ACM=Relative (joystick).

Figure 6 analyzes time and number of errors of each user, using the four control modes. U1 and U2 were faster in performing the exercises, and also made a smaller number of errors. This result was expected, since they are more familiar with laparoscopy and also had some previous experience with laparoscopy simulators. From the histograms, it looks like there is no single control mode preferred by all users, however we can still highlight some basic trends.

Concerning time, it looks like the *T_A* control mode allows all users to perform the exercise in the shortest time. The reason is that *T_A* is the most intuitive way to control the needle, since there is a direct mapping of both position and angles, and is equivalent to moving one's hand in free space. Note, however, that U1 spends most of the time in t_{off} state, trying to get back to where he made a mistake. If we compare times of the users U1 and U2, who are a little trained in laparoscopy, it can be seen that, apart from *T_R*, they score quite similar with the other control modes. In particular U1 (the more trained in the usage of our

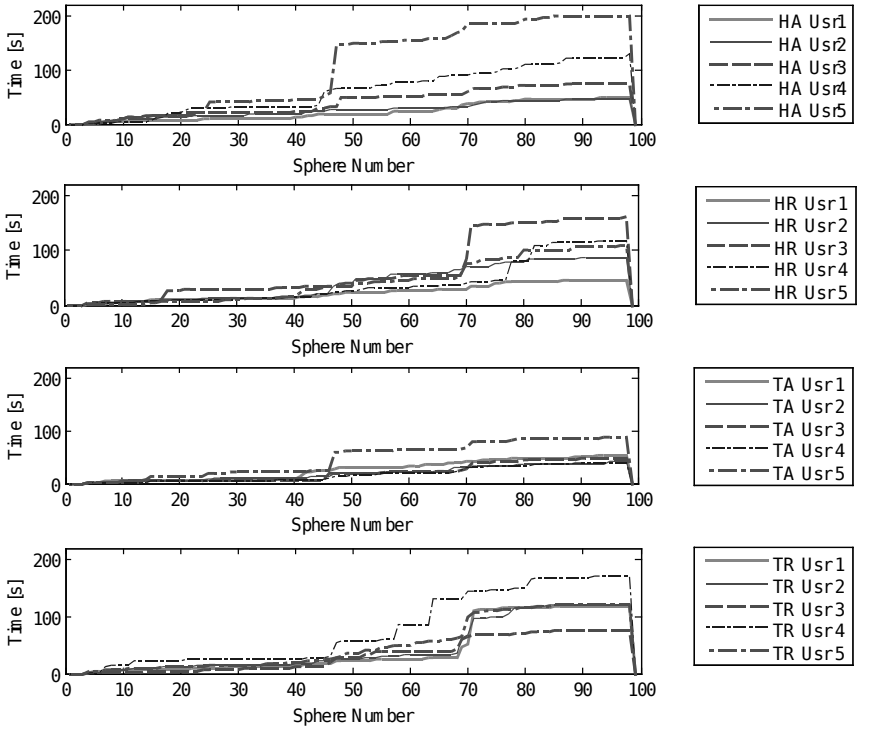


Fig. 5. Plots showing the time spent to complete the exercise by the users with the four control modes. Steep parts highlight difficult points.

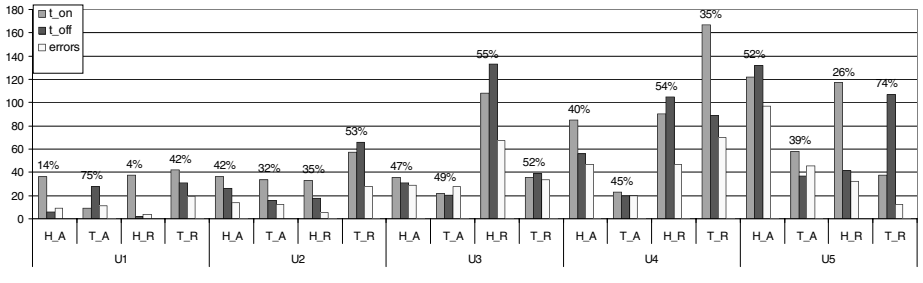


Fig. 6. Histogram showing, for each user and control mode, t_{on}, t_{off} and the number of errors (see text for the description). The percentage shown is the ratio $t_{\text{off}}/t_{\text{total}}$.

hand-held robot), performs as well with *H_R* than with *T_A*, but spending much less time in t_{off} state. U2 has similar performance with *H_R* than with *T_A*.

Concerning number of errors, *H_R* always is the 1st or 2nd choice of 4 out of 5 users. In particular, U1 and U2 make less errors with *H_R* than with *T_A*.

Concerning mean velocity $\bar{v}$, mean acceleration $\bar{a}$ and normalized jerk, trends are less evident. It appears as $\bar{v}_{\text{on}}$ and $\bar{a}_{\text{on}}$ are less when PCM=Handle, suggesting

that the trocar contributes to slow down, but also to control, the movements. Normalized jerk is higher when ACM=Relative, but this is easily explained since the movements of the joystick are, exactly, *in jerks*.

4 Conclusions

We presented a teleoperation platform to assess the performance of different control modes in laparoscopic surgery. The platform is based on a real master, with custom handles mounted on a Phantom haptic interface, and a simulated 6-DoFs slave. This platform has been used to measure performance of users, in order to select the most suitable control mode for a lightweight hand-held robot for laparoscopy that we are developing.

Results show that a direct mapping of position and angles (PCM=Tip, ACM=Absolute), like in the da Vinci system, is the most intuitive for users with no experience in laparoscopy and allows them to complete the exercises faster. However, users which have some training in laparoscopy, performed as fast with our prototype hand-held robot (PCM=Handle, ACM=Relative) and also made a smaller number of errors. This result encourages us to continue the development of the hand-held robot.

Future work will concern the assessment of the best control modes involving also expert surgeons. We expect them to perform well with the hand-held robot, since users U1 and U2 with only a little training in laparoscopy, scored very well. Moreover, we will implement additional exercises, requiring to perform dexterous movements in more challenging conditions.

References

1. Guthart, G.S., Salisbury, J.K.: The Intuitive^{TM} telesurgery system: overview and application. In: Proc. IEEE International Conference on Robotics and Automation – ICRA. (2000) 618–621
2. Focacci, F., Piccigallo, M., Tonet, O., Megali, G., Dario, P.: Design of a hand-held robotic instrument for dexterous laparoscopic surgery. Minim Invas Ther **14**(4-5) (2005) 308–9
3. Lai, F., Howe, R.D.: Evaluating control modes for constrained robotic surgery. In: Proc. IEEE International Conference on Robotics and Automation – ICRA. (2000) 603–609
4. Burdea, G.: Force and touch feedback for virtual reality. John Wiley & Sons, Inc. (1996)
5. Hannaford, B.: Experimental measurements for specification of surgical mechanisms and understanding of surgical skill. Lecture notes, Lecture notes of the European summer school on surgical robotics, Montpellier, France (2003)
6. Cotin, S., Stylopoulos, N., Ottensmeyer, M., Neumann, P., Rattner, D., Dawson, S.: Metrics for laparoscopic skills trainers: the weakest link! In Dohi, T., Kikinis, R., eds.: Proc. Medical Image Computing and Computer-Assisted Intervention – MICCAI. Volume 2488 of Lecture Notes in Computer Science., Berlin, Springer (2002) 35–43

"Virtual Touch": An Efficient Registration Method for Catheter Navigation in Left Atrium

Hua Zhong[1], Takeo Kanade[1], and David Schwartzman[2]

[1] Computer Science Department, Carnegie Mellon University, USA
[2] University of Pittsburgh Medical Center, USA

Abstract. In this paper, we present a left atrium registration system which utilizes a 3D intra-cardiac ultrasound catheter for *faster* (more than 700 times) and higher quality surface registration point collection than current systems and eventually improves the registration *accuracy* and *stability*. With better registration our system can greatly improve the ablation catheter navigation system which is being used in many hospitals to guide left atrium endocardium ablation procedure.

1 Introduction

Recent years, researchers have developed image guided left atrium ablation systems which can automatically register the intra-operative heart shape data to the 3D or 4D model from pre-operative CT scans [1] [2] as well as commercially available Carto Merge system. These systems enable clinicians to navigate the catheter inside left atrium by seeing where the catheter is on computer monitor in real time. Such systems can greatly reduced the difficulty of left atrium ablation procedure, shorten the procedure time and improve its quality. The accuracy of such image guided system is mostly based on the accuracy of its registration. The registration needs clinicians to collect a few intra-operative registration points which clinicians know are on the left atrium endocardium. Then these registration points are used by an ICP [3] algorithm to find a transformation to the heart model coordinate system so that all the points sit on the surface of a left atrium model reconstructed from pre-operative CT scan. If we assume the heart model from CT scan is accurate, the qualify of the registration points determines the quality of registration and thus the accuracy of the whole navigation system.

Under current protocol of commercially available catheter navigation system Carto Merge, clinicians need to manually move a catheter with position sensor to physically touch the endocardium to capture a registration point, which is slow and tedious. The touch is verified by fluoroscopy images. As we know, fluoroscopy imaging device can not be turned on all the time, instead, it can take just a few snapshots. And in fluoroscopy images, heart wall is quite blurred and it makes clinicians hard to decide if the catheter is touching the heart wall. This uncertainty eventually affects the quality of registration points. Now some clinicians are using intra-cardiac ultrasound catheters to guide ablation catheters

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 437–444, 2006.

and verify the touch of ablation catheter to the heart wall [4] [5]. In this case, in addition to the ablation catheter, another ultrasound catheter is inserted into the left atrium. From the real time ultrasound images captured by the ultrasound catheter, clinicians can clearly see the ablation catheter touching the heart wall. The verification of touching is much better than with fluoroscopy images. This is the best scenario currently clinicians can have. But it doesn't improve the speed. To capture 50-60 registration points, a highly experienced clinician may need approximately 10 minutes. Usually in order to capture points which are well spread on the heart wall, more points are needed.

In this paper, we present a novel technique named "virtual touch" which employs an intra-cardiac 3D ultrasound catheter to collect high quality intra-operative surface data at a speed of 700 times faster than current method. The key idea is, with "virtual touch", we capture every surface point we can see in an ultrasound image without physically touching it. When we move the ultrasound catheter and sweep the inner heart wall with ultrasound image plane, we can easily collect thousands of points on the heart wall which give the details of heart shape. It's like a 3D laser scan only replacing laser with ultrasound. The quality of points captured by "virtual touch" is as good as if not better than those captured by physically touching the surface under the guide of ultrasound. More important, with thousands of surface points covering large portion of the left atrium, which is impossible for current systems, the registration's stability has been greatly improved. We will show in our paper that with our "virtual touch", the registration algorithm always find the correct registration independent of initial alignment.

In section 2, we will explain the technique in details. Results will be shown in section 3 with discussions. Finally we will conclude in section 4.

2 Method

The key idea of "virtual touch" is to "see" surface points in ultrasound image to capture them other than to move to and physically touch those surface points to capture them. To do this, our system first automatically detects inner heart wall pixels in ultrasound images and then reconstructs the 3D coordinates of these pixels. We will go through these steps one by one in the following sections.

2.1 Heart Wall Pixel Detection

When parameters of ultrasound machine are tuned accordingly, the heart wall can be seen clearly in ultrasound images. We first use canny edge detector [6] to detect edge pixels. We usually set the thresholds of canny edge detector to a high level in order to collect only wall pixels with high confidence (clear edge). Because our whole algorithm is so fast, we can afford discarding points with even a little ambiguity. This is important to ensure the quality of our registration points.

Sometimes, other chambers of the heart are visible in the far side of the ultrasound image. To avoid including edge pixels of other chambers, we search

along a bunch of beams sent out from the ultrasound image center as the red arrows shown in Figure 1 (a). When we search along one beam and hit one edge pixel, we accept it as a wall pixel, stop searching along the beam and switch to the next beam. Such searching strategy allows us to quickly include all the pixels that are on the wall of the chamber in which the catheter currently is and avoid most wall pixels belongs to other chamber, as shown in Figure 1 (c). More sophisticated contour detection methods could be used here. But we found our technique works just fine and it is very fast.

(a) (b) (c)

Fig. 1. Detect wall pixels: (a) the input image, red arrows sent out from the catheter center represent the direction we search for edge pixels: for each direction, we start from center and when we meet the first edge pixel, we will stop searching along that direction and move to the next direction. (b) the edge detected by canny edge detector. (c) Detection result: red crosses represent the accepted wall pixels. By searching along the red arrows, we can avoid edges from other chambers of the heart.

2.2 Intracardiac 3D Ultrasound

After we detected heart wall pixels in an ultrasound image, we need to reconstruct their 3D coordinates. In our prototype system, we attached Ascension technology's Microbird tracking sensor onto a Siemens's Acuson intra-cardiac ultrasound catheter to build a 3D intracardiac ultrasound catheter. We use single plane phantom [7] to calibrate the 3D ultrasound catheter in a water tank. Temperature corrections has been done to all ultrasound images based on the speed of sound in water [8]. The calibration finds a matrix T and pixel size S_x, S_y, so that the 3D coordinates of any pixel *p(x,y)* in an ultrasound image can be calculated by:

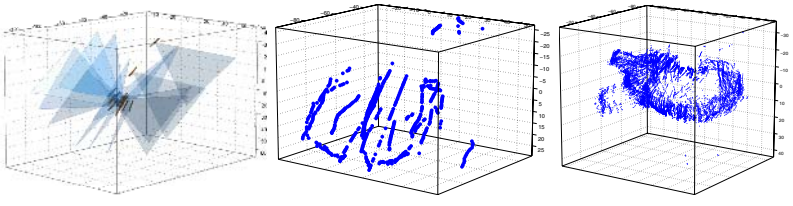
$$P = M_i T \begin{pmatrix} xS_x \\ yS_y \\ 0 \\ 1 \end{pmatrix} \tag{1}$$

where M_i is the direct reading from the magnetic position sensor at the moment the ultrasound image is captured. It's a transformation matrix which gives the position and orientation of the position sensor attached to the ultrasound catheter. S_x, S_y is the size of ultrasound image pixel's width and hight in millimeter. T transforms the 2D pixel coordinate to the 3D coordinate of the position

sensor and M_i transforms from the position sensor's coordinate to the transmitter's coordinate which is fixed during an operation. And eventually we get the 3D coordinate P.

2.3 "Virtual Touch"

As we can automatically detect heart wall pixels in ultrasound images and reconstruct the 3D coordinates of the wall pixels, the surface registration point collection procedure becomes much easier: Clinicians only need to navigate the 3D ultrasound catheter and make the image plane sweep the wall of left atrium. Our system will automatically capture the video output of the ultrasound catheter and detect wall pixels. The 3D coordinates of those pixels can be immediately calculated using equation 1 as shown in Figure 2. No physical touch is necessary: as long as the ultrasound image plane "touches" the heart wall. That's why we call this technique a "virtual touch".



(a) (b) (c)

Fig. 2. Virtual Touch: (a) Clinicians only need to move and rotate the 3D ultrasound catheter so that the ultrasound image plane can sweep large portion of the left atrium. (b) The system can automatically reconstruct the 3D wall points seen by ultrasound images. (c) An example of the heart wall points captured by our system. There are total 12781 points from 427 images.

After we collected registration points, an ICP algorithm will register the points to the left atrium model extracted from pre-operative CT scan. The registration will find a transformation from the magnetic tracking system's transmitter's coordinate to left atrium model's coordinate so that the distance from all registration points to the surface of the left atrium model is minimized. After registration, the ablation catheter which is tracked using the same magnetic transmitter can be visualized with CT left atrium model in real time to help clinicians to navigate it. When synchronized with ECG signal, "Virtual Touch" can also get 4D registration points for system as [2].

3 Results and Discussion

We tested the prototype system with phantom models built from real patient's CT scan. The phantom model's inside cavity has the same shape as the patient's

left atrium. During the test the model was submerged in a water tank. We inserted the 3D ultrasound catheter inside the model and captured some wall points. The points then is registered to the 3D surface model.

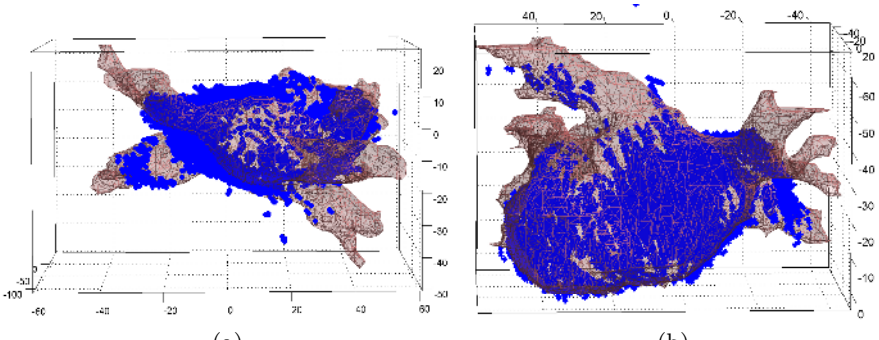
3.1 Registration Speed

During the test, we captured 427 ultrasound images within 3 minutes (at 2.5 frames/second) as shown in Figure 2 (c). Image processing (wall pixel extraction and 3D reconstruction) can be done in the same rate. And after 3 minutes, we successfully reconstructed 12871 surface registration points (heart wall points). Comparing to currently move-and-touch protocol which usually needs about 10 minutes for 50-60 points, our method is more than 700 times faster.

As ICP algorithm only takes seconds to find the registration using currently available workstations, the majority of time for registration process is spent to collect registration point. With 700 times faster, the whole registration process time will be greatly shortened. In our case, with 12871 surface registration points, the ICP algorithm still can be done around 2 minute. So the increase of time used by ICP to process more points is negligible comparing to the time saved to collect these points.

3.2 Registration Quality

We registered the surface points we captured by "virtual touch" to the 3D model from CT. The average distance from points to surface model after registration is 1.2143mm.



(a) (b)

Fig. 3. Registration Result: average distance to surface model is 1.2143mm

As we have stated in section 1 the registration quality is directly determined by the quality of registration points. With registration error around 1mm, we can see the points are of high quality. It is reasonable because under the best scenario, clinicians use ultrasound images to verify the touch of catheter to heart wall to collect surface registration points. Since the left atrium wall is

not rigid, a touch may actually push the wall away from its original position. With our method, no physical touch happens so the wall points we captured will be as good as if not better than what clinicians can have using current technology.

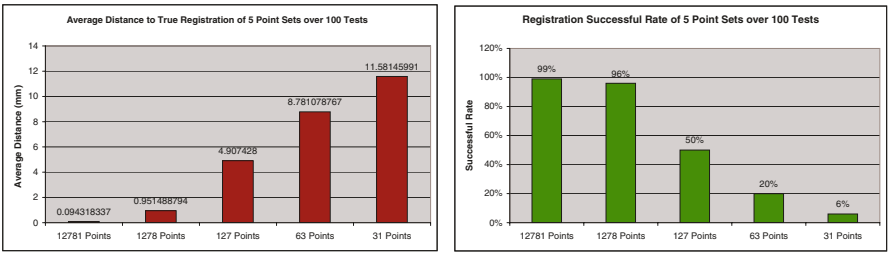
As we can see in Figure 3, there are some outliers in our point set. Those are because of the speckle noise in ultrasound images. We expect there are tens of such outliers exist in our point data set. But given the total number of points in our point set is 12781, these less than 1% outliers can not divert the registration from the correct alignment. And with trimmed ICP algorithm [9], the effect of such outliers can be easily fixed.

3.3 Registration Stability

Besides accuracy, registration stability is also a very important criterion for any registration system. Accuracy tells how precise a system can be and stability tells how likely the system can achieve such accuracy.

All left atrium registration systems actually sample the intra-operative heart shape (surface points) and register this sampled surface shape to pre-operative full shape model (CT). Because the intra-operative sampled heart shape only have partial information, the loss of information will introduce ambiguity for registration. A good registration system can greatly reduce such ambiguity so that it has good chance to find true registration. David Simon [10] showed us given a shape, how to determine the best sample point set to maximize the stability and accuracy. But in reality, before registration we don't know where our catheter is inside heart. Then we can't move the catheter to sample the shape at specified location. Instead, we can only random sample the shape. With only a few samples of the shape, registration may have a lot ambiguity or local minimums. As we add more random samples, we can reduce ambiguity or local minimums so that registration algorithm can have a better chance to converge to the true registration. Or we can say, adding more samples we can increase the registration's stability. Now we would like to know the ***lower bound of*** how many samples we need to ensure a acceptable stability.

We designed a stability test with our phantom data. We generated 100 random initial alignments: ± 30 degree of rotation along x, y and z axis, and ± 5 millimeter translation from ground truth registration. And we randomly sub-sampled our 12781 surface registration points down to four other point sets each having 1278, 127, 63 and 31 points. These points are well spread over the shape. We use these five sets of sample points to do registration from all 100 initial alignments. After each registration, we calculate the average distance of points from this registration to the ground truth registration and then the average distance of all 100 tests. Result is shown in Figure 4 (a). To show how likely each point set will guide the registration to true alignment, we also calculated the registration successful rate over 100 tests. If the average distance from registered points to its location in true registration is less than 1mm, we think the current registration is successful. Result is shown in Figure 4 (b).



(a) (b)

Fig. 4. Stability Test Result: (a) Average distance to true registration of 5 point sets. (b) Successful rate of registration with 5 point sets. If average distance of each registered point to its location in true registration is less than 1mm, we think it is successful.

As we can see, to achieve a reasonable stability, we need to random sample more than 1000 points over the left atrium surface. This number is much higher than any current available registration systems which suggest a sample set of around 50 - 60 (Carto Merge) - 150 (some doctors) points. As shown in our test result, with such few samples of the shape, registration only has a successful rate of 50% or less. The reason all these systems ignore such stability issue is it is too time-consuming (may need hours) for doctors to physically touch 1000 locations on heart wall with catheter. But with "virtual touch", to capture ten thousand points will not take 3 minutes. This ensures a 99% chance to get a true registration with only a rough initial alignment. With such superior stability, clinicians can be much more confident about the registration found with "virtual touch" while with currently available commercial systems, doctors must face the fact that the registration only has a 50% chance to be correct.

4 Conclusion and Future Work

In this paper, we present a novel fully automatic left atrium registration method using "virtual touch". It is *fast*, *accurate* and *stable*. In our phantom model experiment, we showed that our prototype system can collect high quality surface registration points more than 700 times faster than current methods. The registration error with our system is around 1.2mm. Also we showed that registration with "virtual touch" can achieve higher stability which means it is not only accurate but also it is always accurate. Besides all those benefits, the system is easy and cheap to build. Comparing to fluoroscopy imaging devices, our prototype system can do the job much better yet cost much less. It can be easily deployed to hospitals with minimum modification of current available devices.

Animal tests of this new technique has been scheduled. In the future, we will develop faster yet more sophisticated automatic heart wall pixel detection algorithms to minimize the effect of speckle noise in ultrasound images. Also a post-process algorithm could help to eliminate outlier points.

Based on this "virtual touch" technique, many before impossible tasks now become possible. Researchers can use our technique to collect thousands of surface points of left atrium and use part of them as registration point set and rest as test set to evaluate new registration algorithms. We realize that because heart is non-rigid and adjacent to other moving organs, the shape of left atrium changes during operation from what it was during CT scan. With "virtual touch", we can collect enough points to capture even the finest topological details and these points could be used to modify high-resolution models from CT scan and result a better navigation system for left atrium ablation procedure as well as in other heart chambers.

References

1. Verma, A., Marrouche, N.F., Natale, A.: Novel method to integrate three-dimensional computed tomographic images of left atrium with real-time electroanatomic mapping. Cardiovasc Electrophysiol **26** (2004) 365–370
2. Zhong, H., Kanade, T., Schwartzman, D.: Sensor guided ablation procedure for left atrium endocardium. In: Proc. of MICCAI 2005. (2005) LNCS 3750: 1–8
3. Besl, P.J., McKay, N.D.: A method for registration of 3-d shapes. In: IEEE Trans. Pattern Analysis and Machine Intelligence. (1992) 14:239–256
4. Schwartzman, D., Bazaz, R., Nosbisch, J.: Catheter ablation to suppress atrial fibrillation: Evolution of technique at a single center. Interventional Cardiac Electrophysiology **9** (2003) 295–300
5. Verma, A., Marrouche, N.F., Natale, A.: Pulmonary vein antrum isolation: Intracardiac echocardiography-guided technique. Cardiovasc Electrophyisol **15** (2004) 1335–1340
6. Canny, F.J.: A computational approach to edge detection. (In: IEEE Trans. Pattern Analysis and Machine Intelligence) 13: 775–790
7. Prager, R.W., Rohling, R.N., Gee, A.H., Berman, L.: Rapid calibration for 3-d freehand ultrasound. In: Ultrasound in Medicine and Biology. (1998) 24(6): 855–869
8. Bilaniuk, N., Wong, G.S.K.: Speed of sound in pure water as a function of temperature. The Journal of the Acoustic Society of America **3** (1993) 1609–1612
9. Chetverikov, D., Stepanov, D.: Robust euclidean alignment of 3d point sets. In: First Hungarian Conference on Computer Graphics and Geometry. (2002) 70–75
10. Simon, D.A., Kanade, T.: Geometric constraint analysis and synthesis: methods for improving shape-based registration accuracy. In: CVRMed. (1997) 181–190

Towards Scarless Surgery: An Endoscopic-Ultrasound Navigation System for Transgastric Access Procedures

Raúl San José Estépar[1], Nicholas Stylopoulos[2], Randy E. Ellis[1], Eigil Samset[1], Carl-Fredrik Westin[1], Christopher Thompson[1], and Kirby Vosburgh[1,2,3]

[1] Brigham and Women's Hospital
[2] Massachusetts General Hospital
[3] Center of Integration of Medicine and Innovative Technology (CIMIT), Boston, MA
{rjosest, ellis, samset, westin}@bwh.harvard.edu,
{nstylopoulos, ccthompson, kvosburgh}@partners.org

Abstract. Scarless surgery is a new and very promising technique that can mark a new era in surgical procedures. We have created and validated a navigation system for endoscopic and transgastric access interventions in *in vivo* pilot studies. The system provides augmented visual feedback and additional contextual information by establishing a correspondence between the real time endoscopic ultrasound image and a preoperative CT volume using rigid registration. The system enhances the operator's ability to interpret the ultrasound image reducing the mental burden used in probe placement. Our analysis shows that rigid registration is accurate enough to help physicians in endoscopic abdominal surgery where, by using preoperative data for context and real-time imaging for targeting, distortions that limit the use of only preoperative data can be overcome.

1 Introduction

For centuries, the peritoneal cavity has been approached through large incisions of the anterior abdominal wall. In the past two decades, the laparoscopic approach has gained wide acceptance because it offers a safe and less invasive alternative: pain and the complications associated to large abdominal incisions are minimized. To further reduce the invasiveness of peritoneal access, the next logical step is to eliminate the incision through the abdominal wall altogether. Rather, natural orifices may provide the entry points for surgical interventions. Recently, several research groups have been able to access the peritoneal cavity through peroral- transgastric (i.e. through a small incision in the gastric wall) and also peranal-transcolonic approaches to perform organ resections in an animal model [1]. This new approach has the potential to replace or augment laparoscopic techniques currently used to treat many diseases. It may be especially beneficial to obese patients or those who have undergone multiple procedures and thus are at risk for adhesions.

Minimally invasive peroral-transgastric and peranal-transcolonic surgery is in its infancy. Before widespread use of these techniques is possible, several technical barriers must be overcome. These include providing the physician with adequate visual feedback, clear indicators of instrument location and orientation and support in the recognition of anatomic structures. The appearance of the abdominal structures

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 445 – 453, 2006.

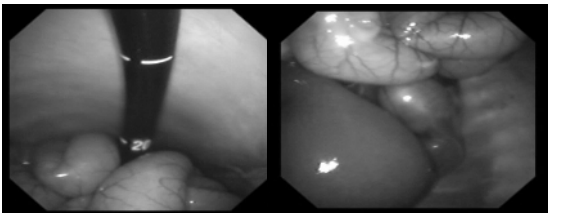


Fig. 1. Endoscopic view of a transgastric procedure

through the transgastric approach is totally different than in an open or laparoscopic approach (see Fig. 1). The unique angle of view, limited light, and the need to insert all instruments through a narrow channel are critical technical challenges. The introduction of the instruments in the peritoneal cavity (e.g. through a gastrotomy) should be performed in a way that precludes damage to surrounding organs and vasculature.

One relatively simple approach to augment the endoscopic/laparoscopic view is to use ultrasound (US). Laparoscopic US (LUS) and endoscopic US (EUS) are often used to guide biopsies and interventional procedures. However, the vast majority of physicians are not comfortable with performing invasive procedures under ultrasound guidance because of the inherent problems of image interpretation.

The navigation of a flexible endoscopic device inside the abdomen brings similar challenges as in traditional laparoscopy, but new complexities are added:

- The flexibility of the endoscope tip makes the understanding of its distal orientation difficult. Unlike laparoscopic procedures, there is no direct observation of the endoscope tip. Because of the lack of a global reference of the tip with respect to the patient body, successful navigation inside the stomach and in the abdomen cavity generally requires the expertise of a highly trained gastroenterologist (up to two years sub-specialization).
- A majority of the structures that are accessible through a transgastric access lie in a retrograde position with respect to the incision in the stomach wall. Access to such locations requires detailed knowledge of the place of the tip with respect to adjacent structures, particularly vessel.
- Understanding the position and orientation of the ultrasound B-scan plane is a ubiquitous problem even for experienced sonographers.
- An inherent problem to any ultrasound based system is the difficulty in interpreting the images because of low contrast, reduced field of view and acoustic windows constrains, despite the close proximity of the US probe to the target organs.

Several groups have attempted to address the orientation and interpretation problem by using preoperative data jointly with the intraoperative US data [2-6]. Lindseth *et al*. have shown that the fusion of intraoperative US images and preoperative MRI enhance the perception by extending the overview of the operating field. Ellsmere *et al*. showed that a 3D display with the main vascular structures and the probe positioning improved the spatial orientation of the operator, thus reducing the time to locate the organ of interest and increasing the operator's certainty.

We present here a system that addresses those challenges and makes intra-cavitary interventional techniques easier to master and use in practice, and thus more likely to be widely adopted. Our system relies on providing *context information* about the interpretation of the US image based on preoperative data (CT or MRI). The system is based on tracking the endoscope tip and US plane with an electromagnetic tracker and established the correspondence of the real time positioning of the instrument tip with respect to preoperative data. The preoperative data is also used to generate 3D models of reference anatomical structure. Those structures are displayed with respect to the position of the probe in real time. An enhance interpretation of the US image is achieved by oblique reformatting the preoperative dataset according to the US plane.

2 Navigation System

The system consists of three major hardware components:

1. A conventional ultrasound machine equipped with a laparoscopic or endoscopic probe.
2. Tracking device comprising a transmitter and receiver sensors.
3. A host computer with a display for the use of the physician.

We use four coordinate systems:

- Instrument coordinate system (*D*): local to the device that is being tracked, namely, e.g. the tip of an endoscopic probe or a surgical pointing device.
- Receiver coordinate system (*R*): local to the sensor mounted on the tracked instrument.
- Transmitter coordinate system (*T*): The coordinate frame of the transmitter of the tracking device, taken as the global coordinate frame.
- Patient coordinate system (*P*): local to the patient in the CT scanner; intraoperatively this is also the coordinate frame of the display.

These coordinate frames are related by affine transformations that are provided either by the tracking system or by computations. Notationally, the transform $\mathbf{T}_{A\rightarrow B}$ is the affine transformation between coordinate system *A* and coordinate system *B*.

2.1 Calibration

Calibration is a necessary and crucial process that is performed to find the transformation $T_{D\rightarrow R}$ between the coordinate system attached to the US B-scan plane and the coordinate system of the position sensor. The real-time position of the B-scan plane is used to generate two of the principal displays (see Fig. 2); this position is unknown until we perform the calibration. We have used the single-wall phantom method as described by Prager *et al.* [7]. The accuracy of the calibration was one of the lower bounds for the final positioning error for the system.

2.2 Registration

The registration step is done intraoperatively with the subject placed on the OR table and before the procedure takes place. The registration transformation $\mathbf{T}_{T\rightarrow P}$ is found in two stages: an initial rigid registration and a real time adaptive registration.

The initial registration is performed by either using anatomical fiducials or high contrast fiducials placed on the skin before the preoperative imaging. Those fiducials are identified in the CT image and in the coordinate system *T* by touching them with a tracked pointer. The registration matrix is solved with Horn's pair-wise point matching method [8]; the resulting transformation is denoted as $\mathbf{T}^{\text{orig}}{}_{T\to P}$.

The real time registration is performed by using an additional position sensor, attached to the subject's thorax, as a local reference frame. This real-time adaptive registration compensates for rigid movements of the patient. The registration transformation $\mathbf{T}_{T\to P}$ is updated according to the expression

$$\mathbf{T}_{T\to P} = \mathbf{T}^{orig}_{T\to P}\,\mathbf{T}^{init}_{R\to T}\left(\mathbf{T}^{m}_{R\to T}\right)^{-1}$$

where $\mathbf{T}^{\text{m}}{}_{R\to T}$ is the transformation given by the sensor attached to the patient and $\mathbf{T}^{\text{init}}{}_{R\to T}$ is the initial transformation reported by the sensor after it is attached to the patient.

2.3 Display

The display consists of three primary elements (see Fig. 2):

- *Display 1*: A 3D scene of the patient skeleton and principal vascular structures obtained from the preoperative dataset and a model of the tracked endoscopic probe and the position and orientation of the EUS plane.

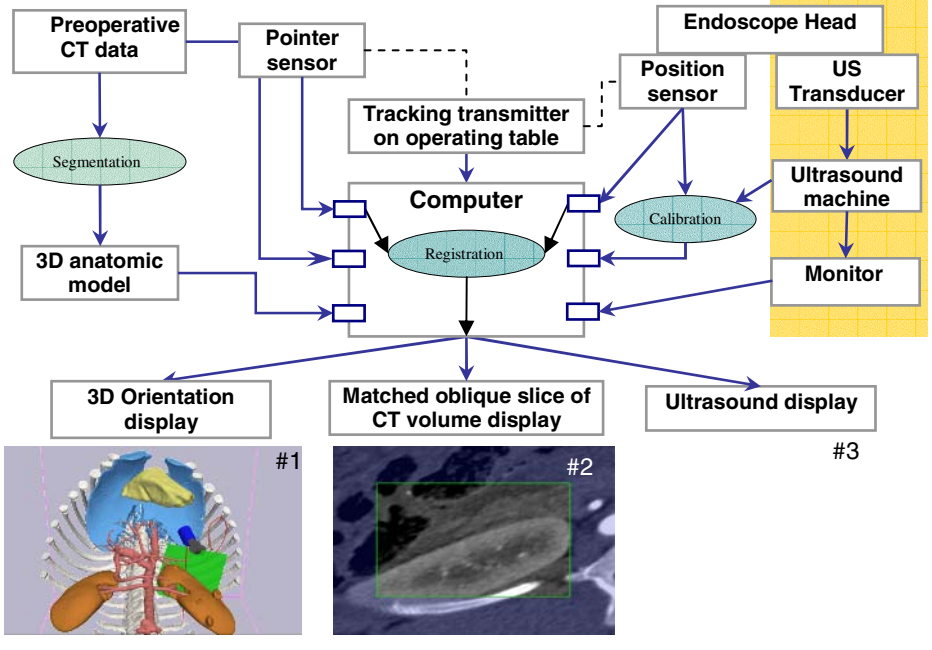


Fig. 2. System description: data flow paths and main displays

- *Display 2*: A reformatted CT image in the oblique plane corresponding to the EUS image. The reformat image is augmented by showing a greater extent that the one covered by the US plane. A square outline shows the area corresponding to the US field of view.
- *Display 3*: Direct acquisition of the EUS image

3 Materials and Methods

The tracking system that we have employed is an electromagnetic tracker (MicroBIRD, Ascension Technology Corp., Burlington, VT) which was connected to the computer through a PCI board. The sensor attached to the endoscope tip was 0.3mm diameter and 1.8 mm length. The attachment of the sensor to the ultrasound probe causes an increase in overall probe diameter of less than 2 mm which is immaterial in the probe use. A similar sensor is used for the real time adaptive rigid registration. The US images were provided by BK Panther Laparoscopic Ultrasound for LUS acquisitions are Olympus EU-C60 for EUS acquisitions. Both ultrasound systems have Doppler capabilities. Preoperative CT were made with a Siemens Sensation 64. Three scans were acquired per study: a baseline scan, a contrast enhanced scan and delayed scan.

Our system was tested in a porcine model under general anesthesia. Free breathing was allowed and forced ventilation was only used during CT scanning to reduce breathing artifacts. Before CT scanning, four high-contrast CT markers were placed on the laterals of the rib cage. The subject was placed on the OR, the fiducial makers were located, and the initial registration was calculated.

To assess the performance of our system, both quantitatively and qualitatively, we ran two set of experiments: one for system error and one for user performance. In all cases we recorded all instrument motions and corresponding US images for retrospective analysis.

3.1 Experiment 1

This experiment, using a tracked LUS probe, was designed to assess the total registration error incurred by our system. Laparoscopic or transgastric procedures normally require insufflation of the abdominal cavity. We hypothesize that the distortion due to insufflation would impose a lower bound for the total registration error for these interventions, whether laparoscopic or trans-gastric.

Four landmarks were used as reference to assess the error. Three landmarks were anatomical features, namely, the branching points between the aorta and three major arteries: the celiac, superior mesenteric (SMA) and right renal. The fourth landmark was a 6 mm diameter bead that was implanted in the left kidney. The bead was clearly observed both in CT and US.

To assess the error, an expert was asked to perform a standard Laparoscopic exploration and to identify several US images where those landmarks were clearly visible. Doppler ultrasound was employed for the vessel landmarks to give an accurate location of the branch point. Independently, the same land-marks were identified in the CT volume. The total registration error was measured as the distance between a CT landmark

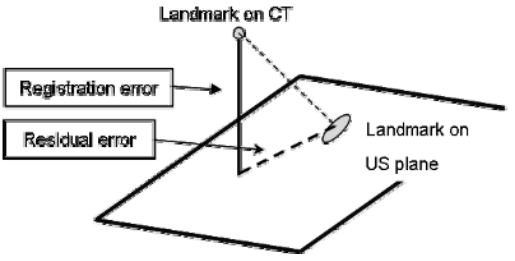


Fig. 3. Registration error definitions

and the US planes where the same landmark was visible (see Fig. 3 - 4b). The position of the US plane in the CT space was known through the transformations given by our system. We considered the error in the normal direction to the US plane to be *primary*, and any error within the US plane to be *residual.* Because our system is intended to provide contextual information, and because the clinician can readily compensate for misalignments between the CT oblique slice and the US plane (as long as the target structure was present in both modalities), the normal error is the limit on how far the clinician needed to search for the target.

We also measured respiratory-induced motion of a target organ by stitching a tracker to the right kidney surface. Tracking data were acquired during free breathing and forced (ventilated) breathing; we recorded the insufflation and the heart rate. Motion data were examined with time-series analysis to find the principal harmonic at that corresponded with the breathing frequency timed during the experiment. After filtering that harmonic in the X, Y and Z time-series, a principal-component analysis (PCA) was performed. The net motion was computed as $2(\lambda_1)^{1/2}$ where λ_1 is the principal eigenvalue of the covariance matrix.

3.2 Experiment 2

A group of 3 experts and 5 novices in US guided endoscopic interventions were requested to localize a predefined list of targets on a fixed time of 5 minutes. The users ran this task twice, using the conventional EUS technique and using our Image Registered Gastroscopic Ultrasound (IRGUS) system. We interleaved the tasks between users, so half used the IRGUS system first and the other half used the EUS first. During the experiment, the location and orientation of the probe was recorded and we noted which structures were properly identified. Novice users were always assisted by an expert. From the positioning data, a kinematic evaluation of the user's motion was performed to characterize performance during the task [9]. Kinematic analysis provides measurements for the amount and smoothness of motion that is required to achieve the task. Finally, a questionnaire was used at the end of the task to assess subjective responses to the navigation system.

4 Results

The total registration error was computed for each landmark using 4 US slices where the landmarks were visible. The average total error is shown in Fig. 4a. The error upper boundary is approximately 5 mm range. Because the landmark features were also of that size, the landmarks should have always been visible in the reformatted CT plane. We confirmed this by retrospectively evaluating the results: reviewing the data and

Table 1. Kinematic analysis comparing our system (IRGUS) and conventional endoscopic approach (EUS)

	Path Length (cm)	*Smoothness of Motion* d^3d/dt^3	*Depth Perception* (cm)	*Response Orientation* (radians)
EUS	1600.3	12.6	9877.5	52.5
IRGUS	1245.7	9.2	8174.2	42.3

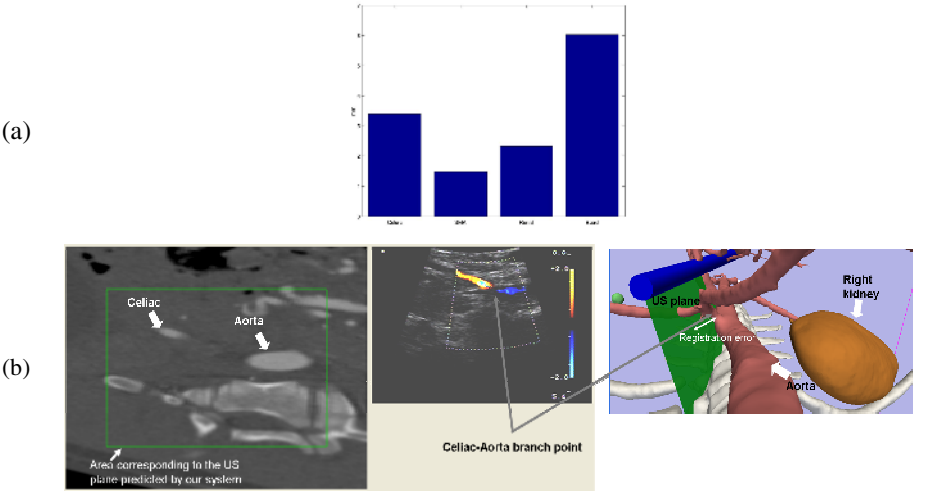


Fig. 4. Registration error for 4 landmarks. (a) Average error for different US planes where the landmarks were visible. (b) Example of one plane used for the evaluation of the error corresponding to the celiac branch.

reconstructing views, we confirmed that all landmarks were visible in both in the US plane and the CT oblique at times when the expert reported localizing the landmarks.

Our data for respiratory-induced kidney motion were acquired during free breathing and ventilation with an insufflation of 200ml. The subject's heart rate was steady at 70 beat/s. The principal harmonic of motion was 0.6 Hz, which corresponds with the breathing frequency that was timed during the experiment. The total displacement in the direction of maximum variance was 1.8mm for forced breathing and 1.4mm for free breathing.

Our second experiment showed that using the conventional EUS, novices identified only 29% of the structures and experts 50% within the allotted time. In contrast, IRGUS significantly increased to 71% and 80% respectively. In addition, the analysis of kinematic data showed that using IRGUS the physicians not only identified more structures, but they were more efficient as well (see Table 1). IRGUS improved efficiency of conventional EUS by 17-27% in the analyzed characteristics. All differences were statistically significant at the level of $p<0.05$.

5 Discussion

The utility of our system does not depend on absolute spatial precision. Our hypothesis is that the orthogonal distance between the US plane and a target on the preoperative data is on an acceptable range (less than 5mm) to allow context information. We also found that when targets appeared in both the US plane and in the reformatted CT – regardless of the amount of displacement within that plane – users found the contextual information very useful in guiding interventions. A major concern was that the motion of organs (such as the kidney) induced by respiration would compromise the utility of our system, but our experiments showed that this motion was limited and was significantly lower than the registration errors

A key component of our approach is that we have relaxed the accuracy requirements for registration of patient's anatomy given by the US to the preoperative volumetric images. This "*reference"* registration consists on relying on an initial rigid registration of the scanner space to the patient space, plus a real time correction of this initial rigid registration computed by tracking the patient position with a sensor. By complementing real time imaging with closely registered preoperative images, we aim to improve the way in which real-time images are interpreted, but without relying on high accuracy registration methods needed by traditional image-based navigation systems [10]. We claim that *reference* registration is particularly suited to endoscopic abdominal surgery where, by using preoperative data for *context* and real-time imaging for *targeting*, distortions that limit the use of only preoperative data can be overcome. We observed that the accuracy of our approach is within surgically acceptable limits and that the contextual information provided by our navigation system improved the performance of both expert and novice users.

Novice clinicians performing US guided endoscopic interventions found the system easy to master and improved their confidence in the identification of different structures thanks to the closely registration to the CT. The number of structures that they were able to correctly identify was double than without the assistance of the system. Even expert clinicians found the navigation system of great help, increasing their confidence in the imaged structured by the real-time US and reducing their intervention time.

Our system can assist in the implementation of transgastric interventions. One of the major difficulties while performing the intervention is to avoid major stomach vessels when puncturing the stomach wall. This puncture can be performed with minimal bleeding is those vessels are avoided. However, if the vessel is accidentally comprise, fatal iatrogenic injury could lead to the death of the patient. Our system can assist in this crucial task by tracking the position of the tip of the instrument in relation to the vessels of the stomach and other abdominal structures.

Acknowledgments

This work has been supported by the US Department of the Army under award DAMD 17-02-2-0006 to CIMIT. The information does not necessarily reflect the position of the government and no official endorsement should be inferred. This work has been partially supported by NIH P41-RR13218. The authors want to acknowledge the material and support provided by the following companies: Ascension Technologies, B-K Ultrasound, and Olympus.

References

1. M.S. Wagh, B.F. Merrifield, C.C. Thompson, "Survival studies after endoscopic transgastric oophorectomy and tubectomy in a porcine model", *Gastrointest. Endosc.*, vol. 63, pp. 473–478, 2006.
2. J. Harms, H. Feussner, M. Baumgartner, et al, "Three-dimensional navigated laparoscopic ultrasonography", *Surg. Endosc*, vol. 15, pp. 1459–1462, 2001.
3. J. Ellsmere, J. Stoll, D. Rattner, D. Brooks, R. Kane, W. Wells, R. Kikinis, and K. Vosburgh, "A Navigation System for Augmenting Laparoscopic Ultrasound", Lectures Notes in Computer Science, vol. 2879, pp. 184–191, 2003.
4. F. Lindseth, S. Ommedal, J. Bang, et al.: "Image Fusion of Ultrasound and MRI as an Aid for assessing Anatomical Shifts and improving overview and interpretation in Ultrasound Guided Neurosurgery". CARS 2001, 1230:247–252, 2001.
5. T. Lange, S. Eulenstein, M. Hunerbein. "Augmenting Intraoperative 3D Ultrsound with Preoperative Models for Navigation in Liver Surgery". *MICCAI 2004 Part II*, Saint Malo, Springer LNCS 3216, C Barilot, DR Haynor, Eds, pp. 26–29, 2004.
6. Becker HD, Herth F, Ernst A, Schwarz Y. "Bronchoscopic Biopsy of Peripheral Lung Lesions Under Electromagnetic Guidance: A Pilot Study", *J Bronchology*, vol. 12, pp. 9–13, 2005.
7. R. Prager, R. Rohling, A. Gee, et al. "Rapid calibration for 3-D freehand ultrasound". Ultrasound Med Biol, vol 24, pp. 855– 869, 1998.
8. B.K.P. Horn, "Closed-form solution of absolute orientation using unit quaternions", Journal of the Optical Society of America A, vol. 4, pp. 629–642. 1987.
9. N. Stylopoulos, S. Cotin, S.K. Maithel, M. Ottensmeyer, P.G. Jackson, R.S. Bardsley, PF Neumann, D.W. Rattner, S.L. Dawson. "Computer-enhanced laparoscopic training system (CELTS): bridging the gap". *Surg Endosc*. Vol 18, pp. 782–789, 2004.
10. G. Penney, J. Blackall, M. Hamady, et al. "Registration of freehand 3D ultrasound and magnetic resonance liver images", Medical Image Analysis, vol. 8, pp. 81–91, 2004.

New 4-D Imaging for Real-Time Intraoperative MRI: Adaptive 4-D Scan

Junichi Tokuda[1], Shigehiro Morikawa[2], Hasnine A. Haque[3], Tetsuji Tsukamoto[3], Kiyoshi Matsumiya[1], Hongen Liao[4], Ken Masamune[1], and Takeyoshi Dohi[1]

[1] Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1, Hongo, Bunkyo, Tokyo 113-8656, Japan
junichi@atre.t.u-tokyo.ac.jp
[2] Biomedical MR Science Center, Shiga University of Medical Science, Seta-tsukinowa Cho, Otsu, Shiga 520-2192, Japan
[3] GE Yokogawa Medical Systems Ltd., 4-7-127, Asahigaoka, Hino, Tokyo 191-8503, Japan
[4] Graduate School of Engineering, The University of Tokyo, 7-3-1, Hongo, Bunkyo, Tokyo 113-8656, Japan

Abstract. Aiming at real-time 3-D visualization of organ motion to navigate surgical procedures in MRI-guided surgery, a new 4-D MR imaging technique called "Adaptive 4-D Scan" has been proposed. The technique is designed to acquire a time series of volumetric 3-D images (4-D image) of cyclically moving organ, even in a low-field open-configuration MR scanner. A pre-operative 4-D image is acquired with respiratory phase parameter, which is monitored by using navigator-echo-based real-time tracking of the liver and diaphragm. During operation, the respiratory phase is again monitored in real-time, and a 3-D image, reflecting the current state of the target organ, is extracted from the pre-operative 4-D image and provided to physicians as a pseudo real-time 3-D image. We implemented Adaptive 4-D Scan into a 0.5 Tesla open-configuration clinical MRI system for intervention. Phantom and volunteer studies were performed to assess feasibility of this technique, in terms of image quality, imaging time and position accuracy of the imaged subject. A 4-D image (matrix: $256 \times 128 \times 10 \times 8$) of cyclically moving phantom was acquired in 719 s, and RMS position error between the imaged subject and the real subject was 2.3 mm, where the range of motion was 50 mm. 4-D image of the moving liver was also successfully acquired under near clinical condition. In conclusion, the study shows that the proposed method is feasible and has capability to provide real-time dynamic 3-D atlas for surgical navigation.

1 Introduction

Intraoperative Magnetic Resonance Imaging (MRI) has been recognized as a promising tool for surgical guidance in recent years. With the introduction of open-configuration MRI, variety of surgical application has been explored to investigate the benefit of intraoperative imaging with soft tissue discrimination.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 454–461, 2006.

Despite the merit of accessible MRI in surgery, its limited applicability to moving organs is unsolved issue. Generally, it takes more than a few seconds to acquire even a slice of MR image; therefore an image inherently has misalignment to the actual organ when the images are provided to the physician. Furthermore, fast motion of the target causes serious artifact on a MR image, which makes it difficult for physicians to discriminate the lesion and make accurate decision in targeting. These are the major issues in intraoperative MRI before expanding its application target to moving organ in abdominal cavity, and have been discussed frequently in the community [1]-[3].

To capture an image of organs moving due to respiration, several techniques have been proposed. Liu et. al. proposed navigator gating technique [4], which enables to reduce motion artifact by accepting MR echo signal data only when the respiratory state is in the certain range (gating window). Although the method has been widely used for image diagnosis of abdominal organs, it is not suitable for surgical guidance on moving organ, because images provide only static information at constant respiratory state. As an extended technique of navigator gating, Kolmogorov demonstrated multiple volume acquisition [5]. This technique allows us to acquire multiple images at different respiratory states, but number of gating windows, which equals to number of respiratory states, is determined by histogram of motion, and thus it is impossible to visualize full range of motion of the subject. Another approach is 4D MR imaging using internal respiratory gating proposed by Siebenthal et. al[3]. In this study, the 3D images of the liver were reconstructed for complete respiratory cycle by retrospective stacking of dynamic 2D images using internal image-based gating. Unfortunately, since the proposed method is based on fast 2-D imaging (in case of their study, they use sequence with frame rate of 175 ms/frame), the type of imaging sequence that can be used in this method is limited. Furthermore, a low-field open-configuration MR scanner, which is thought to be an ideal guidance tool for surgery, usually doesn't have capability to drive such a fast imaging sequence.

Based on these backgrounds, we propose a new 4-D MR imaging method, called Adaptive 4-D Scan, which can be implemented to a low-field open-configuration MR scanner, and is applicable for real-time 3-D surgical navigation. The ultimate goal of this work is to generate pseudo real-time 3-D image during operation, through following steps: (Step 1) a time series of 3-D images of a target organ covering whole respiratory cycle is acquired before operation using proposed method, (Step 2) a real-time tracking of the target organ performed during operation to detect respiratory phase of a patient, and (Step 3) best-fitted 3-D image is selected at each moment during operation based on the real-time respiratory phase parameter, and a pseudo real-time 3-D image is provided to physicians.

In this paper, we evaluated 4-D images in terms of image quality as well as position error between actual subject and imaged subject. To the authors' knowledge, none of the related work have argued image position error with a scope to guide surgery. However this is important issue for real-time surgical guidance, since the position information is used to guide surgical incision

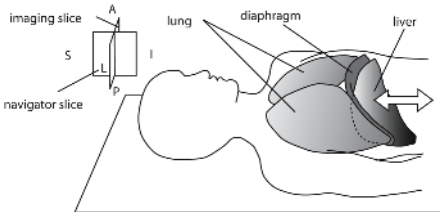


Fig. 1. To monitor respiratory phase of the patient, the shift of the liver was detected by acquiring navigator echoes from the sagittal slice. The liver mainly shifts in S-I direction with diaphragm during breath. Images were acquired from axial slices.

e.g. needle insertion, ablation. In addition, the liver of healthy volunteer was scanned with proposed method to assess feasibility on near clinical conditions.

2 Materials and Methods

Adaptive 4-D Scan technique is similar to respiratory gating, which enables to acquire single 3-D image of cyclically moving subject. The major difference between these two methods is adaptive update of scan parameters based on real-time feedback of respiratory phase to the pulse sequencer.

In this section, we describe two key techniques to implement Adaptive 4-D Scan: real-time respiratory monitoring to detect respiratory phase during scan, and adaptive control of scan parameters for 4-D imaging.

2.1 Real-Time Respiratory Monitoring Using MR

The basic principle of real-time respiratory monitoring is based on the navigator echo technique, originally proposed by Ehman[6]. A navigator echo is a magnetic resonance echo to detect the motion of the subject during image acquisition. Since only one or a few navigator echoes are required to detect the motion, time-resolution of motion tracking is much higher even than 2-D MR imaging.

In this study, we employed a navigator echo to detect the respiratory phase of the subject in real-time, by tracking shift of the liver, that moves mostly along superior-to-inferior axis during respiration as shown in Fig. 1. To acquire imaging echoes and navigator echoes simultaneously, we configured a pulse sequence, in which navigator echo acquisitions are inserted into repetition of imaging echo acquisitions. N_{intv}, number of imaging echo acquisitions between navigator echo acquisitions, is configurable. Frame rate of respiratory monitoring is determined as $\frac{1}{TR \times N_{\mathrm{intv}}}$ fps, where TR is repetition time of MR echo acquisitions. In this study, we set $N_{\mathrm{intv}} = 5$.

2.2 Adaptive Control of Scan Parameters for 4-D Scan

To reconstruct a 4-D image, $k_x \times k_y \times z \times t$ matrix needs to be filled up with MR echo signals $S(k_x, k_y, z, t)$, where k_x and k_y are frequency and phase encoding parameters for each, z is the slice offset and t is time. t can be replaced by a respiratory phase parameter ϕ, since we assume that motion of the target organ is cyclic. We note the number of steps of ϕ as N_ϕ, which equals to image

resolution in time dimension. In contrast to normal 3-D imaging, where k_y and z can be determined prior to start of a scan, the scan parameters are dynamically updated based on the respiratory phase ϕ of the subject in Adaptive 4-D Scan. This is because ϕ reflects the state of patient's free breath, and its variation is unpredictable before the scan.

To obtain ϕ, we detect the shift of the liver and diaphragm in superior-to-inferior direction by using navigator echo. In this study, ϕ was defined as $\phi = \frac{s}{L}\pi$, where s is the shift of the liver and diaphragm from the reference point, and L is maximum stroke of s. Note that the reference point is at the end of the stroke, and thus the range of s is $0 \geq s \geq L$.

The example pseudocode for adaptive control of scan parameters would be:

```
repeat begin
  φ = get_resp_phase();
  h = get_volume_index(φ);
  i = z_index_table[h];
  j = ky_index_table[h];
  acquire_signal(z_i, k_{y_j})
until is_scan_complete(z_index_table, ky_index_table)
```

where `z_index_table[]` and `ky_index_table[]` are tables to preserve z and k_y for each volume indexed by h. In procedure `get_resp_phase()`, a pulse sequence to acquire navigator echo is executed, and ϕ is returned. `get_volume_index()` converts ϕ to the volume index h. To convert ϕ, which is a continuous value, to the index h, the range of variation of ϕ is split into N_ϕ partitions (Fig. 2). The width of each partition is determined in two ways: (A) $\Delta\phi_h = const.$ and (B) $T_h = const.$ for $h = 1, 2, ..., N_\phi$, where $\Delta\phi_h$ is width of each partition of ϕ and T_h is total duration while ϕ is in h^{th} partition. The merit using (B) is that echo signals are distributed to each partition evenly, and therefore imaging time can be reduced without compromising N_ϕ. We implemented this method to an open-configuration MR system, GE Signa SP/i 0.5 T (GE Healthcare, Chalfont St. Giles, UK).

3 Experiments and Results

3.1 Phantom Study

We conducted the phantom study to evaluate image quality, imaging time and position accuracy of 4-D images, which are important issues for image guidance.

To simulate motion of a human liver, a cylindrical agar phantom was placed on a specially designed stage, which generated reciprocating motion with stroke of $L = 50$ mm (Fig. 3). The crank was rotated manually with speed of approx. 0.2 cycles per second.

In this study, the phase of the cyclical motion ϕ was determined by the shift of the phantom itself. To monitor ϕ, the navigator echoes were acquired from a

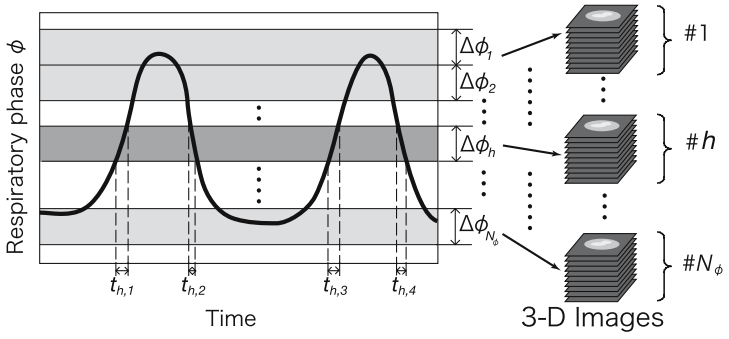


Fig. 2. The range of respiratory phase variation is split into N_ϕ partitions. Two methods to determine width of each partition were implemented: (A) $\Delta\phi_1 = \Delta\phi_2 = ... = \Delta\phi_{N_\phi}$, where $\Delta\phi_h$ is width of the h^{th} partition. (B) $T_1 = T_2 = ... = T_{N_\phi}$, where h^{th} is amount duration while ϕ is in the h^{th} partition. T_h is defined by $T_2 = t_{2,1} + t_{2,2} + t_{2,3} + t_{2,4} + ...$ in the graph.

plane parallel to the motion of the phantom. Imaging planes were also in the same direction, to measure the shift of the phantom from each 2-D slice image in 3-D images independently. Consequently, the sample number of the measurements for one 4-D image equals to the number of slices.

The pulse sequence was based on fast gradient echo (FGRE), with parameters: TR:30 ms, TE:9.7 ms, flip angle:30 degrees and FOV:300×300 mm. The matrix size of 4-D images were $256 \times 128 \times 10 \times N_\phi$. 4-D images were acquired with $N_\phi = 4, 6, 8, 10$ using partition method (A), and with $N_\phi = 6$ using partition method (B). Scans were performed twice for each condition.

We assessed imaging time, image quality and position error of imaged phantom for each 4-D image. "Ground-truth" for error evaluation was calculated based on the translation of the table. The table position was calibrated to the coordinate of the scanner. Since each 3-D image was reconstructed from MR echo signals acquired with different ϕ, distributed within the corresponding phase partition, we defined "ground-truth" position for each 3-D image at the center of each partition. "Ground-truth" shift of the phantom from the base position (at the end of the stroke) in h^{th} image can be calculated by $s_h = \frac{L}{\pi}(\frac{\Delta\phi_h}{2} + \sum_{k=1}^{h-1} \Delta\phi_k)$.

The part of result 4-D images for $N_\phi = 4, 6, 8$ and 10 acquired using partition method (A) are shown in Fig.4, and the root mean square (RMS) error of phantom position and imaging time are shown in Fig. 5 (1). We found that there are trade-offs between image quality and imaging time, and between image position accuracy and imaging time except the case $N_\phi = 10$.

In addition, the images acquired using partition method (A) and (B) are compared fixing N_ϕ at 6. Fig.5 (2) shows the relation between image position error and distance from the center of the stroke. In both cases, position errors are higher at the centers of the stroke than at the ends of the stroke, but difference between minimum and maximum errors is much larger in method (A). The

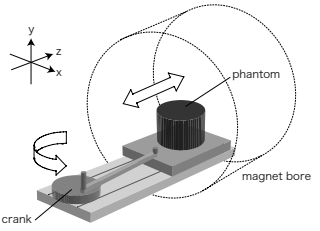


Fig. 3. The phantom was placed on the stage with a crank mechanism to make reciprocating motion with stroke of 50mm. The motion of the stage was limited along only one direction.

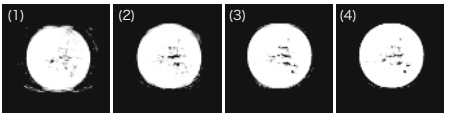


Fig. 4. 2-D images as part of result 4-D images are shown: (1) $N_\phi = 4$, (2) $N_\phi = 6$, (3) $N_\phi = 8$ and (4) $N_\phi = 10$. Image quality was improved when N_ϕ was set to larger value.

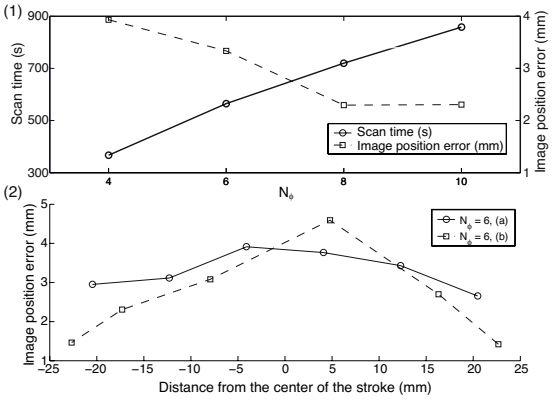


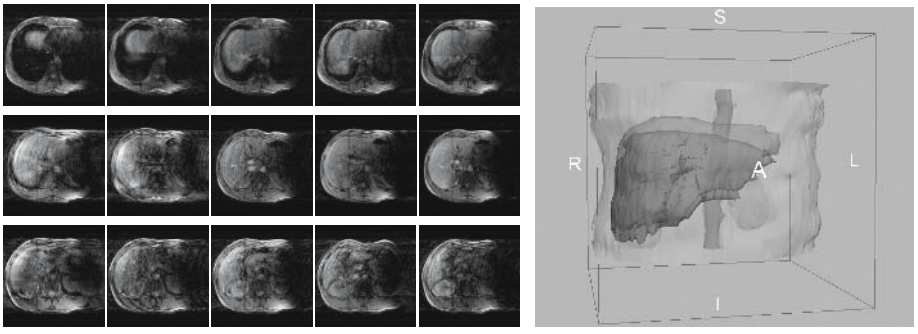
Fig. 5. Errors and time evaluation of the phantom study. RMS position errors and imaging time for $N_\phi = 4, 6, 8$ and 10 fixing partition method to (A) are shown in (1). Also relation between errors and position of the phantom in method (A) and (B) are compared fixing N_ϕ in (2). Note that two 4-D images containing 10 slices were acquired for each condition, and errors were evaluated for each slice independently. Therefore the number of samples to calculate RMS error is 20. Imaging times are average time of two trials.

imaging time for (A) and (B) were 565 s and 472 s, and time reduction by 16 percents was achieved.

3.2 Volunteer Study

To evaluate feasibility of Adaptive 4-D Scan, we performed study on the volunteer's liver. 4-D image of the liver of the healthy volunteer was scanned on near clinical conditions, while the volunteer was freely breathing. Adaptive 4-D Scan was performed with $N_\phi = 5$, and number of slices of 21, using partition method (A). The pulse sequence was FGRE (TR:28 ms, TE:8.7 ms and flip angle:30 degrees). The scan slices were axial for imaging.

The 4-D image of the liver was successfully acquired in 1807 s, and the part of the 4-D image is shown in Fig.6(1). The stroke of diaphragm shift in S-I direction was 13.9 mm. The liver in each 3-D volume was manually segmented and a surface model was generated (Fig. 6(2)).



(1) (2)

Fig. 6. 4-D image sized $256 \times 128 \times 21 \times 5$ pixels was acquired, in the volunteer study. As a part of the 4-D image, three slices of volunteer's liver at five different respiratory phases are shown in (1). Organs, vascular and other atlas can be discriminated. 3-D surface models of the liver at two different respiratory phases were created from the 4-D image, visualized on the surgical navigation software, 3D Slicer [7] in (2). The skin, aorta and kidneys are also visualized as position references.

4 Discussion

We demonstrated the new method to acquire 4-D image of moving organ with a scope to generate pseudo real-time 3-D image during operation by selecting best-fitted 3-D image from 4-D image based on real-time respiratory phase monitoring. The phantom study showed that RMS position error of imaged phantom was 2.3 mm in case that the stroke of the motion was 50 mm and N_ϕ was 8. The error is clinically acceptable level, since the typical size of tumors for treatment is about tens millimeters. Image quality and imaging time were also evaluated changing N_ϕ, and we found that there are trade-offs between imaging time and image quality or position accuracy. Imaging, however, time can be reduced by changing phase partitioning method, and we succeeded to reduce imaging time by 16 percents in the phantom study. The volunteer study demonstrated that the proposed method can be applied to the real human liver. Although the imaging time of approx. 30 minutes is thought to be applicable, a same evaluation on a patient under general anesthesia, breathing with respirator, need to be performed, since histogram of motion of the target is different from free-breath.

There are two factors that control image position accuracy and image quality in Adaptive 4-D Scan: width of respiratory phase partition, $\Delta\phi$, and time lag of real-time respiratory phase monitoring and adaptive control. $\Delta\phi$ effects inconsistency in position of the phantom, where MR echo signals are acquired for each 3-D image. This inconsistency causes position error and blur/noise in images, as shown in the phantom study. The second factor, time lag, leads inaccuracy in monitoring of ϕ and also causes inconsistency in position of the phantom. Time lag was determined by repetition time (TR) of echo signal acquisition and throughput of the scanner system, and currently it is more than 100 ms. In Fig.5 (1), position error is improved up to $N_\phi = 8$, but it is almost unchanged when

$N_\phi > 8$. We think this is because the effect of time lag is constant for all N_ϕ, while the effect of $\Delta\phi$ can be reduced by increasing N_ϕ, and primary factor controlling the position error is changed from $\Delta\phi$ to time lag around $N_\phi = 8$.

The frame rate of pseudo 3-D navigation image generated from 4-D image is theoretically limited by that of tracking and, more than 30 fps could be achieved by using the developed system. We think that the proposed method would be suitable for automated targeting on moving organ for MRI-guided robotics [8] and focused ultrasound surgery (FUS)[9].

In conclusion, we proposed Adaptive 4-D Scan technique for real-time intraoperative MRI. Phantom and volunteer study shows the proposed method is feasible for clinical use and has capability to provide real-time dynamic 3-D atlas for surgical navigation.

This work is partially supported by Research Fellowships of the Japan Society for the Promotion of Science (JSPS) for Young Scientists.

References

1. Blackall, J. M., King, A. P., Penney, G. P., Adam, A., Hawkes, D. J.: A statistical model of respiratory motion and deformation of the liver. Medical Image Computing and Computer-Assisted Intervention - MICCAI 2001, LNCS 2208 (2001) 1338-1340
2. Tokuda, J., Hirano, M., Tsukamoto, T., Dohi, T., Hata, N.: Integration of Projection Profile Matching into Clinical MR Scanner System for Real-time Organ Tracking and Image Registration, Medical Image Computing and Computer-Assisted Intervention - MICCAI 2003, LNCS 2879 (2003) 311–318
3. von Siebenthal, M., Cattin, P., Gamper, U., Lomax, A.,Székely G.: 4D MR Imaging Using Internal Respiratory Gating, Medical Image Computing and Computer-Assisted Intervention - MICCAI 2003, LNCS 2879 (2003) 311–318
4. Liu, Y.L., Riederer, S.J., Rossman, P.J., Grimm, R.G., Debbins, J.P., Ehman, R.L.: A monitoring, feedback, and triggering system for reproducible breath-hold MR imaging, Magnetic Resonance in Medicin **30** (1993) 507–511
5. Kolmogorov, V.N., Watts, R., Prince, M.R., Zabih, R., Wang, Y.: Simultaneous multiple volume (SMV) acquisition algorithm for real-time navigator gating, Magnetic Resonance Imaging **21** (2003) 969–975
6. Ehman, R. L., Felmlee, J. P.: Adaptive Technique for High Definition-MR Imaging of Moving Structures Navigator Echoes. Radiology **173** (1989) 255–263
7. Gering, D. T., Nabavi, A., Kikinis, R., Hata, N., O'Donnell, J., Wells III, W. M.: An Integrated Visualization System for Surgical Planning and Guidance Using Image Fusion and an Open MR. J. of Magn. Reson. Im. **13** (2001) 967–975
8. Chinzei, K., Hata, N., Jolesz, F. A., Kikinis, R.: MR compatible surgical assist robot: System integration and preliminary feasibility study. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2000, LNCS 1935 (2000)921–930
9. Jolesz, F.A., Hynynen, K.: Magnetic resonance image-guided focused ultrasound surgery. Cancer J., **8** (2002) S100–S112

The Use of Super Resolution in Robotic Assisted Minimally Invasive Surgery

Mirna Lerotic and Guang-Zhong Yang

Royal Society/Wolfson Foundation Medical Image Computing Laboratory,
Imperial College of Science, Technology and Medicine, London SW7 2BZ, UK
Mirna.Lerotic@imperal.ac.uk, g.z.yang@imperial.ac.uk

Abstract. In minimally invasive surgery, a small field-of-view is often required for achieving a large magnification factor during micro-scale tasks such as coronary anastomosis. Constant change of the orientation and focal length of the laparoscope camera, however, is cumbersome and can impose extra visual and cognitive load to the operating surgeon in realigning the visual pathways and anatomical landmarks. The purpose of this paper is to investigate the use of fixational movements for robotic assisted minimal invasive surgery such that the perceived resolution of the foveal field-of-view is greater than the intrinsic resolution of the laparoscope camera. The proposed technique is based on super resolution imaging using projection onto convex sets. Validation with both phantom and *in vivo* data from totally endoscopic coronary artery bypass surgery is provided.

1 Introduction

The use of computer assisted technology in surgical applications has risen dramatically in recent years, contributing to a range of new methods for training, education and diagnosis. In surgery, advances in medical image computing have permitted detailed preoperative planning and intra-operative surgical guidance. One of the most promising advances in surgical technology in recent years is the introduction of robotic assisted Minimally Invasive Surgery (MIS) [1]. It is increasingly being used to perform procedures that are otherwise prohibited by the confines of the operating environment. The technique offers a unique opportunity for deploying sophisticated surgical tools that can greatly enhance the manual dexterities of the operating surgeon. The future of robotic surgery is in the intelligent use of preoperative data for more complex procedures such as beating heart surgery where there is large-scale tissue deformation.

Whilst the benefit of patient recovery and surgical outcome is well established, the deployment of MIS is associated with the complexity of instrument controls, restricted vision and mobility, difficult hand-eye co-ordination, and the lack of tactile perception which require a high degree of operator dexterity [2]. For robotic assisted MIS, dexterity is enhanced with microprocessor controlled mechanical wrists, which allow motion scaling for reducing gross hand movements and performing micro-scale

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 462 – 469, 2006.

tasks that are otherwise not possible. The issue of resolving restricted vision, however, has so far achieved limited progress. This is largely due to the restriction on the number of trocar ports used during MIS and the small field of view (FOV) required for achieving a large magnification factor for performing micro-scale tasks such as coronary anastomosis. The consequence of restricted vision significantly affects the visual-spatial orientation of the surgeon and the awareness of the peripheral sites. During MIS, constant change of the orientation and focal length of the laparoscope camera is cumbersome and can impose extra visual and cognitive load to the operating surgeon in realigning the visual pathways and anatomical landmarks. In certain cases, this can also contribute to surgical errors as view rotation is known to affect surgical performances. The limited FOV has also made intra-operative surgical guidance difficult due to the paucity of unique surface features for aligning pre- or intra-operative 3D image data.

Recently, gaze contingent robotic control has attracted significant research interest due to its unique capability in coupling human visual perception with machine vision [3, 4]. The research is based on the fact that in terms of visual acuity, human eyes do not have a uniform visual response, in fact, the best visual acuity is only within a visual angle of one to two degrees. This is called foveal vision, and for areas that we do not direct our eyes towards when observing a scene, we have to rely on a cruder representation of the objects offered by non-foveal vision, of which the visual acuity drops off dramatically from the centre of focus. The limited extent of the fovea demands the eyes being highly mobile and able to sweep across a large visual angle. As a result, the motion of the eyes is extremely varied in both the amplitude and frequency spectrum. During the periods of visual fixation, small eye movements continuously move the projection of the image on the retina. These fixational eye movements include small saccades, slow drifts, and physiological nystagmus. Existing research has shown that micro-motions, including both micro-saccades of the eyes [5] and subpixel (sub-sampling) movements of visual scene [6], can enhance visual resolution, although the mechanisms of the human visual system are unknown.

The purpose of this paper is to investigate the use of fixational movements to robotic assisted MIS such that the perceived resolution of the foveal FOV is greater than the intrinsic resolution of the laparoscope camera. In practice, this permits the use of a relatively large FOV for micro-surgical tasks such that the aforementioned drawbacks associated with the existing MIS setup can be avoided. With the proposed method, the position of the surgeon's gaze can be located by using binocular eye tracking and super resolution can then be used to provide a high resolution image to the foveal FOV. The proposed technique is based on super resolution imaging using projection onto convex sets (POCS) [7, 8] and validation with both phantom and *in vivo* data from totally endoscopic coronary artery bypass (TECAB) surgery is provided.

2 Method

Super resolution is a method used to get a high resolution image from a set of low resolution images that have sub-pixel shifts [6]. The additional information required

to calculate the high resolution image is provided by different "views" given by sub-pixel shifts of low resolution images, and therefore provides a way of pushing the image resolution beyond the intrinsic limitation of the image sensors.

If one considers high resolution (HR) image $\mathbf{x}$ to be the true representation of the scene, then a low resolution (LR) image $\mathbf{y}_i$ is equal to

$$\mathbf{y}_i = \mathbf{DMH}_i\mathbf{x} + \mathbf{v}_i \tag{1}$$

where $\mathbf{D}$ is a down-sampling operator, $\mathbf{M}$ is a warp operator which accounts for translation, rotation and other transformations , $\mathbf{H}$ incorporates the point spread function (PSF) of the sensor such as motion blur, and $\mathbf{v}_i$ describes the additive noise. The recovery of the HR image $\mathbf{x}$ from a set of sub-pixel shifted LR images $\mathbf{y}_i$ is the goal of all super-resolution algorithms. Existing approaches to the problem include non-uniform interpolation (which consists of sub-pixel LR image registration, interpolation to the HR grid and image restoration), constrained least squares (CLS) [9], and maximum a posteriori (MAP) [10].

The super-resolution reconstruction algorithm used in this work is based on the method of projection onto convex sets (POCS) [7, 8]. The POCS method is an iterative method which incorporates prior knowledge by restricting the solution to the intersection of closed convex sets. In each iteration, a solution $\mathbf{x}^{n+1}$ is derived such that

$$\mathbf{x}^{n+1} = P_m P_{m-1} \ldots P_2 P_1 \mathbf{x}^n \tag{2}$$

where P_i are projection operators which project the arbitrary solution $\mathbf{x}^n$ onto m closed convex sets. Each convex set represents a constraint on the solution such as amplitude, energy, or reference-image constraint [8]. In other words, the projection operators impose prior knowledge about the HR image on the solution.

Tekalp *et al* [7] introduced the following closed convex set constraint for each pixel of each low resolution frame $\mathbf{y}_i$

$$C_i = \left\{ \mathbf{x}_{k,l} : \left| \mathbf{r}_{i;m,n}^{(\mathbf{x})} \right| \le \delta_0 \right\}, \tag{3}$$

where δ_0 defines a bound that represents the statistical confidence in observed LR frames, and is usually estimated from the noise present in the LR images. The $\mathbf{r}_i^{(\mathbf{x})}$ in a sense represents the difference between the HR image $\mathbf{x}$ convolved with a PSF and the *i*-th LR image $\mathbf{y}_i$

$$\mathbf{r}_{i;m,n}^{(\mathbf{x})} = \mathbf{y}_{i;m,n} - \sum_{k,l} \mathbf{x}_{k,l}\mathbf{h}_{i;\,m,n;k,l} \tag{4}$$

where $\mathbf{h}_i$ is a PSF of the sensor for the *i*-th frame, given by the ratio of the area where LR pixel (m,n) and HR pixel (k,l) overlap divided by the total area of the LR pixel. The projection of a HR image $\mathbf{x}$ onto convex set C_i defined in Eq. (3) is given by [7]

$$\mathbf{x}_{k,l}^{n+1} = \mathbf{x}_{k,l}^{n} + \begin{cases} \dfrac{r_{i;m,n}^{(x)} \mp \delta_0}{\sum_{o,p} \mathbf{h}_{i;m,n;o,p}^2} \mathbf{h}_{i;m,n;k,l} & if \ \mathbf{r}_{i;m,n}^{(x)} \gtrless \pm \delta_0 \\ 0 & if \ -\delta_0 \leq \mathbf{r}_{i;m,n}^{(x)} \leq \delta_0 \end{cases} \tag{5}$$

In this study, the reconstruction of the HR images involves the use of the projection operator presented in Eq. (5) together with amplitude and energy constraints. For super-resolution imaging, accurate image registration is crucial to the subsequent reconstruction result, and for this reason the method proposed by Keren *et al* [11] was used.

In order to validate the proposed super-resolution framework for MIS, a Stäubli RX60 robotic arm was used to control the sub-pixel movement of the camera. The system has six degrees of freedom (DOF) and a repeatability accuracy of ±0.02mm at high speed and acceleration. A phantom heart model was created by using thixotropic silicone mould rubber and pre-vulcanised natural rubber latex with rubber mask grease paint to achieve a specular appearance and high visual fidelity. A video sequence was captured using a camera mounted on the described robotic arm. The camera was moved in a zig-zag pattern with the total movement spanning over few pixels. From the video sequence, 40 images were extracted and each image was separately down-sampled to one fourth of the original resolution, to ensure that the algorithm remains "blind" to individual frame shifts. To assess the *in vivo* applicability of the technique, a video sequence from a totally endoscopic coronary artery bypass surgery was used. The image shifts in this case were provided by the natural sub-pixel movement of the camera, present it the video.

3 Results

Fig. 1 shows the example images for the phantom heart experiment, where Fig. 1(a) shows a video snap-shot of the "true", actual scene (original image), Fig. 1(b) is the

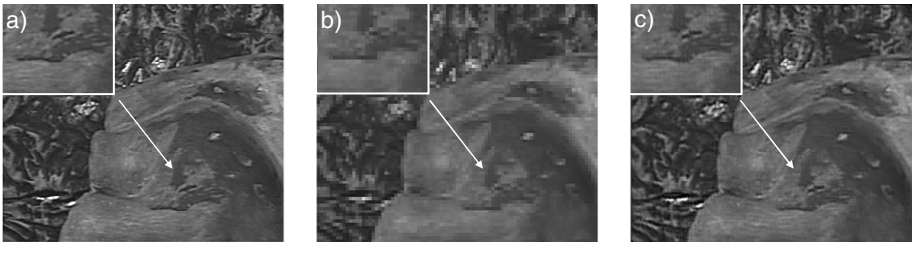


Fig. 1. (a) A video snap-shot of the actual scene (original image) and the corresponding LR images obtained by down-sampling (b). (c) The resulting image after the application of the described POCS super-resolution algorithm, where the insert demonstrates the high resolution details recovered by the proposed algorithm.

corresponding LR image obtained by down-sampling, and Fig. 1(c) illustrates the result of the described POCS super-resolution algorithm as applied to the set of 40 registered LR images. The insert of each figure shows a magnified view of the FOV pointed by the arrow, demonstrating the high resolution-details that have been recovered by the proposed POCS algorithm.

To quantify the level of detail reconstructed in the HR image, an entropy measure [12] that represents the amount of the information present in the images is calculated. Larger entropy values are in general indicative of the high resolution details recovered. For the phantom study, the entropy value of the Original, LR, and POCS reconstructed HR images are 6.74, 6.57, and 6.72, respectively, which suggests the information recovered by the HR image is close to 99.7% based on the entropy measure. It should be noted, however, that this is a global statistical measure, and may not represent the actual visual fidelity recovered. Nevertheless, it does provide a quantitative index when combined with visual feature assessment.

In order to examine the convergence behavior of the algorithm, Fig. 2 (a) illustrates the RMS error in relation to the iteration steps. The algorithm in general has a rapid convergence rate, which makes it suitable for real-time implementations. In Fig. 2(b-c), the power spectra of the LR and HR images are also provided, which demonstrate the amount of high-frequency details recovered.

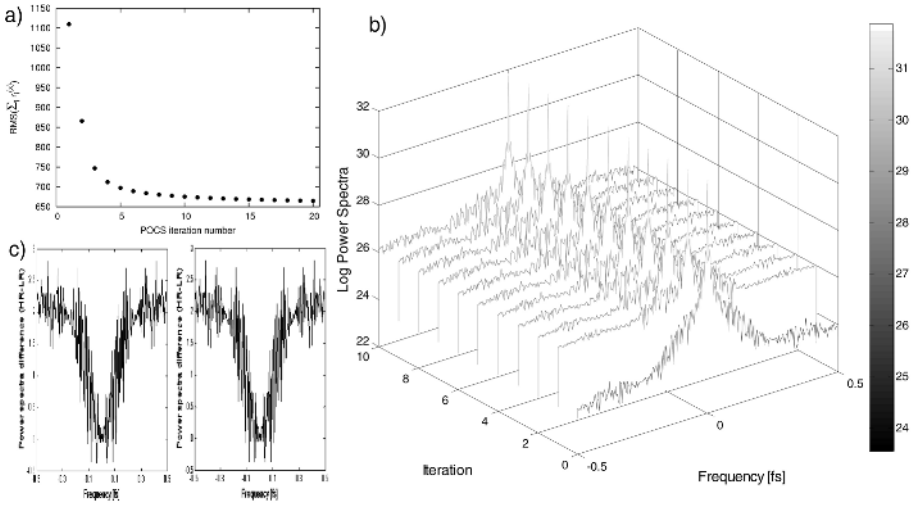


Fig. 2. (a) The convergence of the POCS algorithm as demonstrated by the RMS sum of $\mathbf{r}_i^{(\mathbf{x})}$ defined in Eq. (4). (b) Power spectra of the LR image (at zero position) and HR images for the first 10 iterations (where fs denotes the spatial sampling frequency). The plot shown in (c) shows a difference of power spectra (HR (10^{th} iteration)-LR).

For *in vivo* data, Fig. 3 illustrates an FOV corresponding to the foveal region of the fixation and the high-resolution details recovered by the POCS algorithm, where the corresponding images before resolution enhancement are provided as a reference. The figure also indicates the position of the FOV in relation to the operating view. To provide a more accurate assessment of the level of details recovered, Fig. 4 illustrates the intensity profiles as marked in Fig. 3 for each of the images presented. The corresponding entropy measures for these images are summarized in Table 1.

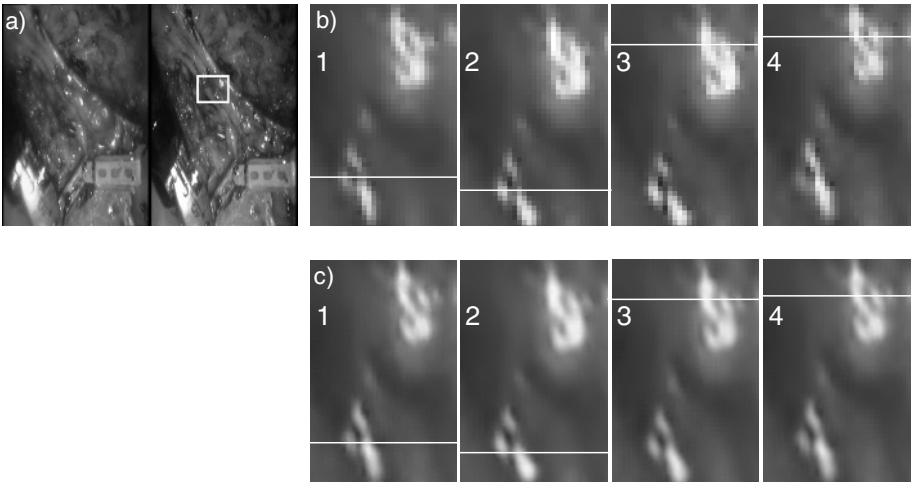


Fig. 3. (a) Image frames from a TECAB sequence; (b) the corresponding magnified time frames from the region marked by the white square in (a); (c) HR reconstructions calculated using the described POCS method. White lines on (b-c) indicate rows at which intensity profiles are shown in Fig. 4.

4 Discussion and Conclusions

In this paper, we have demonstrated the use of super resolution for robotic assisted minimally invasive surgery. This described method presents a novel approach to add an artificial zoom in MIS. The basic motivation of the technique is to investigate the use of fixational movements to robotic assisted MIS such that the perceived resolution of the foveal FOV is greater than the intrinsic resolution of the laparoscope camera. This will allow the use of a relatively large FOV for micro-surgical tasks such that super-resolution is applied to fixation points in response to real-time eye tracking. Experiments with both phantom data under controlled environment and time series from a TECAB surgery were used, and the results derived by using POCS reconstruction demonstrate the achieved improvement in spatial resolution. In combination with eye tracking, the described method can yield gaze contingent super-resolution foveal vision to the surgeon, thus minimizing the need to move the laparoscope camera during surgery.

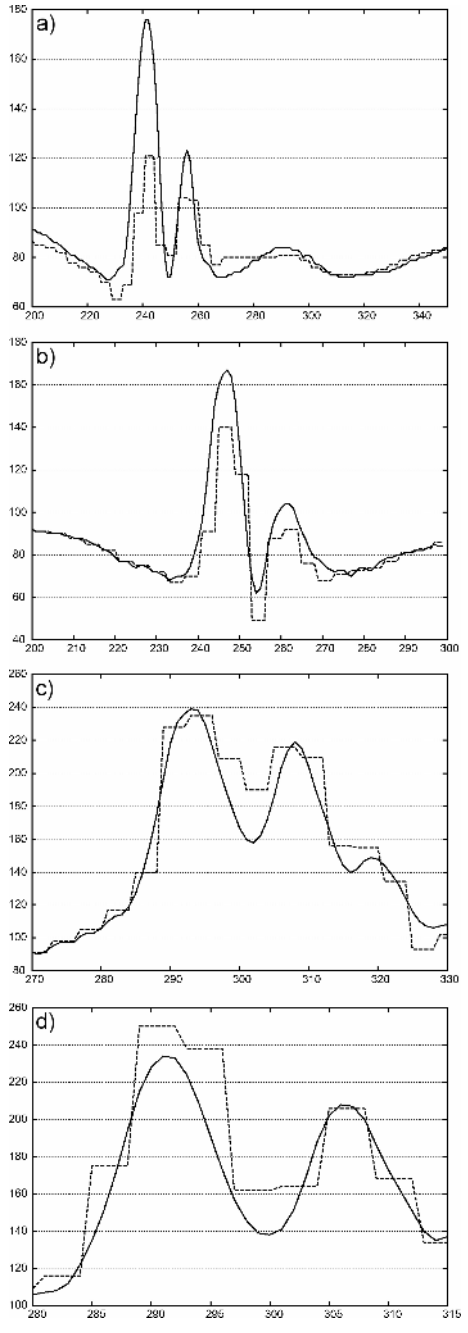


Fig. 4. (a)-(d) Pixel intensity profiles across image features on both HR (solid lines) and LR (dashed lines) images showing the improved resolution of the reconstructed image for the time series 1-4 shown in Fig. 3. Pixels used for line profiles are marked by white lines in Fig. 3(b-c).

Table 1. Entropy values for the TECAB images shown in Figs. 3(b-c)

Time series	LR image	HR image	Difference HR - LR
1	6.72	6.83	0.11
2	6.78	6.82	0.04
3	6.79	6.85	0.06
4	6.79	6.84	0.05

Acknowledgements

The authors would like to thank George P. Mylonas for the help in recording the video sequence of the phantom, and RCUK Basic Technology Programme for financial support.

References

1. Ballantyne, G.H., *Robotic surgery, telerobotic surgery, telepresence, and telementoring.* Surgical Endoscopy, 2002. **16**(10): p. 1389-1402.
2. Tendick, F., et al., *Sensing and Manipulation Problems in Endoscopic Surgery: Experiment, Analysis, and Observation.* Presence, 1993. **2**(1): p. 66-81.
3. Mylonas, G.P., A. Darzi, and G.-Z. Yang, *Gaze Contingent Depth Recovery and Motion Stabilisation for Minimally Invasive Robotic Surgery*. Lecture Notes in Computer Science. 2004. 311-319.
4. Mylonas, G.P., et al., *Gaze-Contingent Soft Tissue Deformation Tracking for Minimally Invasive Robotic Surgery.* Lecture Notes in Computer Science. 2005. 843-850.
5. Martinez-Conde, S., S.L. Macknik, and D.H. Hubel, *The role of fixational eye movements in visual perception.* Nature Reviews Neuroscience 2004. **5**(3): p. 229-40.
6. Sung Cheol, P., P. Min Kyu, and K. Moon Gi, *Super-resolution image reconstruction: a technical overview.* Signal Processing Magazine, IEEE, 2003. **20**(3): p. 21-36.
7. Tekalp, A.M., M.K. Ozkan, and M.I. Sezan. *High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration.* in *IEEE International Conference on Acoustics, Speech, and Signal Processing,* ICASSP-92. 1992, p. 169-172
8. Stark, H. and P. Oskoui, *High-resolution image recovery from image-plane arrays, using convex projections.* J. Opt. Soc. Am. A, 1989. **6**(11): p. 1715-1726.
9. Hong, M.-C., M.G. Kang, and A.K. Katsaggelos. *An Iterative Weighted Regularized Algorithm for Improving the Resolution of Video Sequences.* in *Proc. 1997 IEEE International Conf. on Image Processing.* 1997, p. 474-477
10. Schultz, R.R. and R.L. Stevenson, *Extraction of high-resolution frames from video sequences.* Image Processing, IEEE Transactions on, 1996. **5**(6): p. 996-1011.
11. Keren, D., S. Peleg, and R. Brada. *Image sequence enhancement using sub-pixel displacements.* in *Proceedings CVPR '88., Computer Society Conference on Computer Vision and Pattern Recognition.* 1988. Ann Arbor, MI, USA, p. 742-746
12. Gonzalez, R.C. and R.E. Woods, *Digital Image Processing*. 2002, New Jersey: Prentice-Hall, Inc.

Modeling the Human Aorta for MR-Driven Real-Time Virtual Endoscopy

Klaus J. Kirchberg[1,2], Andreas Wimmer[2], and Christine H. Lorenz[1]

[1] Siemens Corporate Research Inc., Princeton, NJ, USA
{Klaus.Kirchberg, Christine.Lorenz}@siemens.com
[2] Chair for Pattern Recognition, University of Erlangen-Nuremberg, Germany
Andreas.Wimmer@informatik.uni-erlangen.de

Abstract. As interventional magnetic resonance imaging (iMRI) is getting closer to clinical practice, new means of visualization and navigation are required. We present an approach to create a virtual endoscopic view inside the human aorta in real-time. In our approach, defined cross-sectional slices are acquired and segmented in a highly optimized fashion. A geometric shape model is fit to the segmentation points and continuously updated during the intervention. The physician can then view and navigate inside the structure to plan the intervention and get immediate feedback about the procedure. As a component of this system, this work focuses on the segmentation of the cross-sectional images and the fitting of the shape model. We present a real-time 2D segmentation implementation for this application domain and a model fitting scheme for a generalized cylinder (GC) model. For the latter we employ a new scheme for choosing the local reference frame.

1 Introduction

Magnetic resonance imaging (MRI) has been receiving growing attention as guidance modality for minimally invasive procedures. Unlike X-ray fluoroscopy or computed tomography (CT), interventional MRI does not expose patient or physician to potentially harmful ionizing radiation. Also MRI shows detailed soft tissue contrast.

Commercially available MR scanners and their user interfaces are mainly tailored to diagnostic tasks. Visualization of scanned image slices and volumes is optimized for visual inspection on a high-definition display. Interventional MRI on the other hand requires real-time presentation of the acquired image data in the most meaningful way for the physician. Different approaches to MR image visualization for interventional procedures have been proposed. Current techniques include 2D image display of single or multiple slices in real-time [1,2], 2D projections of magnetic resonance angiography (MRA) data [3] and 3D volume rendering [4].

As off-line diagnosis tool, *virtual endoscopy* has been investigated over the last 10 years and found its way into clinical practice, e.g. for the detection of polyps in the colon. In [5], the authors propose *real-time virtual endoscopy* as a new navigation tool for MR-guided interventions. In contrast to diagnostic virtual endoscopy,

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 470–477, 2006.

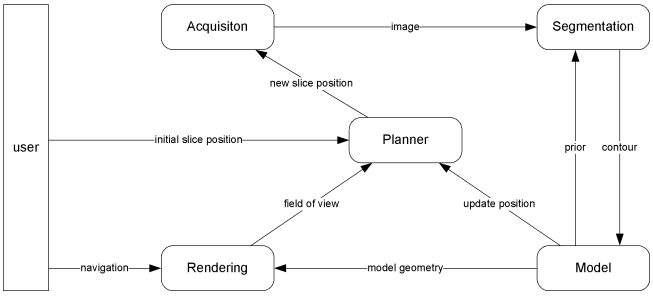


Fig. 1. Data flow in the real-time virtual endoscopy system

the 3D view reflects changes of the anatomy in real-time. The view is continuously updated using on-line MR acquisition, segmentation, and modeling.

In this work we focus on techniques for MR-guided interventions in the human aorta. Coarctation (narrowing) and aneurysms (enlargements) are examples of aortic defects that can be treated by minimally invasive catheter based procedures under X-ray or MRI guidance. During the treatment, a stent graft is inserted and deployed at the appropriate position. We hypothesize that 3D monitoring methods can help improve the accuracy of stent placement as well as reduce the chance of accidental tissue rupture, and improve early detection of complications.

For real-time interaction with the MR scanner we used a communication framework as described in [5]. It provides means of changing slice position and orientation during scanning as well as sending the resulting image data to the processing system.

The planned workflow is as follows. Image slices are acquired perpendicular to the aorta's main axis. The scanned images are first subject to segmentation. Here, points on the aortic wall are extracted and their 3D position is determined by applying the (known) transformation of the image slice. After segmentation, the 3D coordinates of the contour points are added to the cloud of data points. An optimization is continuously fitting and updating the aorta model to best match the point cloud. The coherence information of contour points within one slice is not necessary in our approach, thus leading to a more general system that can easily be extended to other organs. Using the current model shape knowledge, new image slices are requested from the scanner. The layout of the system is depicted in figure 1.

Therefore, as a component of such a real time interventional system, we present methods for (1) optimized segmentation of the aorta and (2) a modified generalized cylinder (GC) model, both suitable for real time application. In particular, this work describes a GC model [6] for the human aorta that is usable in the real-time MRI scenario. A novel reference frame scheme is introduced to solve the problem of singularities in the Frenet-Serret frame. The presented approach is kept general, making it easily adaptable to other organs.

We also present a highly optimized segmentation using Geodesic Active Contours suitable for real-time MRI.

2 Method

2.1 Segmentation of the Aortic Cross-section

Oriented roughly perpendicular to the aorta's main axis, the image slices of the healthy aorta exhibit a more or less circular cross-section. In case of a pathology, e.g. an aneurysm, the contour can diverge even to a concave shape. To handle this variability, segmentation is performed using a level set based approach.

The interventional scenario imposes severe constraints on acquisition and processing time. Feedback of the intervention should be less than a second. To achieve reasonable update detail the system should be able to acquire and process at least 5 frames per second with an in-plane resolution of about 2 mm.

Description of the segmentation method. We segment the aortic cross-section using Geodesic Active Contours [7]. An initial curve C^0 (see below) is placed in the image and deformed by a curvature dependent speed κ and a constant speed ν in curve normal direction $\mathbf{n}$ according to equation (1). g is inversely related to the edge strength which is calculated from the input intensities I. It acts as a stopping function and slows down or stops the curve evolution at object boundaries. The second part of the right hand side in equation (1) constitutes an edge attraction speed in normal direction which increases the robustness for partially weak boundaries or boundaries with small gaps [7], a highly desirable property for images with reduced quality due to rapid acquisition times.

$$\partial_t C = g(\kappa + \nu)\mathbf{n} - (\nabla g \cdot \mathbf{n})\mathbf{n} \tag{1}$$

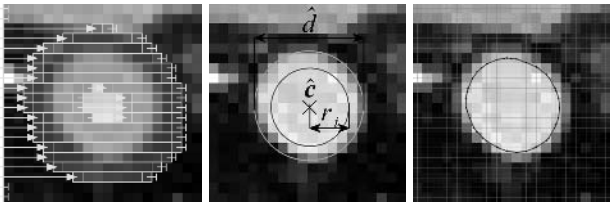
We use the Perona-Malik diffusivity [8] given by equation (2) as stopping function g. λ is a contrast parameter which controls the decrease of the stopping function with increasing edge strength. The gradient is approximated using the Sobel Operator.

$$g(\mathbf{x}) = \frac{1}{1 + \left(\frac{|\nabla I(\mathbf{x})|}{\lambda}\right)^2}. \tag{2}$$

The curve evolution defined in equation (1) is tracked by the level set framework according to equation (3) [7]. The curve is embedded as the zero level set of the 2D level set function ϕ. In contrast to an explicit parametric representation, this implicit representation does not require any (re-)parameterizations during the evolution.

$$\partial_t \phi = |\nabla \phi| \left(\operatorname{div} \left(g \frac{\nabla \phi}{|\nabla \phi|} \right) + \nu g \right). \tag{3}$$

We use a semi-implicit numerical implementation of the level set PDE. This permits large time steps and thus only few updates of the level set function are required until the object boundaries are reached by the deforming curve.



(a) (b) (c)

Fig. 2. (a) Linked list structure to maintain the narrow band rows (b) Placement of the initial curve C^0 (c) Extraction of the final curve. See text for details.

A semi-implicit formulation usually comes with the burden of a high computational effort per update. In [9], a discretization of equation (3) is proposed which allows the application of the Additive Operator Splitting technique [10]. Thereby the computational effort is drastically reduced, and an overall speedup of an order of magnitude compared to an explicit formulation of the level set PDE is reported.

The Narrow Band approach [11] is used to further increase the efficiency of the segmentation algorithm. The computational domain is restricted to grid points in a neighborhood around the zero level set, thus reducing the complexity from $O(N^2)$ to $O(N)$ for a $N \times N$ grid. We use the linked list structure proposed in [12] to maintain which grid points are part of the narrow band. Figure 2 (a) shows the structure describing the narrow band rows. A modified Chamfer distance transformation [13] is used to re-initialize the narrow band once the zero level set comes close to its boundary.

Placement of the initial curve. The initial curve C^0 is placed within the cross-section and expanded in order to recover the boundary. Given estimates $\hat{\mathbf{c}}$ and $\hat{d}$ on the center and approximate diameter of the cross-section from our model of the aorta, the initial curve is set to a circle around $\hat{\mathbf{c}}$, as illustrated in figure 2 (b). The radius r_i is set to a smaller value than half the diameter $\hat{d}$ in order to ensure that the initial curve is within the cross-section also for inaccurate estimates.

Extraction of the final curve. The level set function takes a steady state when the deforming curve has reached the object boundaries. The final curve is then extracted in sub-grid accuracy using the Marching Squares algorithm (see figure 2 (c)).

2.2 Aorta Model

Geometric shape models and their application in medical image analysis have been studied extensively. A survey of methods can be found in [14].

A common deformable model for cylindrical structures is the generalized cylinder model [6]. A generalized cylinder consists of a 3D space curve, called spine or centerline, and a variable 2D cross-section function. The modeled surface is then defined by sweeping the cross-section function along the spine using some sweep

rules. However, this formulation has several weaknesses, e.g. the occurrence of discontinuities at certain points. To overcome those problems, several extensions have been proposed, e.g. [15,16].

We employ a circular cross-section function. This scheme models the healthy aorta as well as aortic abnormalities such as coarctation or aneurysm reasonably well while providing numerical stability for the model optimization process detailed below. In order to fit the model to the segmented surface points we define a cost functional that measures the quality of the fit. We then derive the Euler-Lagrange equation and perform the optimization by gradient descent.

Equation 4 depicts the cost functional, where $\mathbf{c} : [0,1] \mapsto \mathbb{R}^3$ defines the centerline curve and $r : [0,1] \mapsto \mathbb{R}$ the radius function of the cross-section parametrized over the same interval. The function $\Gamma_\sigma : \mathbb{R} \mapsto \mathbb{R}$ denotes the gaussian kernel weighting the influence of data points. Parameter $\hat{s}_i$ refers to the projection of data point $\mathbf{p}_i$ onto the centerline curve as shown below. The fixed start and end points of the curve are $\mathbf{x}_0$ and $\mathbf{x}_1$. The cost functional is regularized by the second and third term, weighted with factors λ and μ, respectively.

$$I(\mathbf{c}, r) = \int_0^1 \frac{1}{N} \sum_{i=1}^{N} \Gamma_\sigma(s - \hat{s}_i)(||\mathbf{c}(s) - \mathbf{p}_i|| - r(s))^2 \, \mathrm{d}s + \tag{4}$$

$$\lambda \int_0^1 \frac{1}{2} ||\nabla \mathbf{c}(s)||^2 \, \mathrm{d}s + \mu \int_0^1 \frac{1}{2} r'(s)^2 \, \mathrm{d}s$$

$$\mathbf{c}(0) = \mathbf{x}_0 \ , \quad \mathbf{c}(1) = \mathbf{x}_1 \ , \quad \hat{s}_i = \arg\min_s ||\mathbf{c}(s) - \mathbf{p}_i|| \tag{5}$$

The corresponding Euler-Lagrange equations give rise to the gradient descent equations

$$\frac{\partial \mathbf{c}}{\partial t}(s) = -\frac{1}{N} \sum_{i=1}^{N} \Gamma_\sigma(s - \hat{s}_i) \, (||\mathbf{c}(s) - \mathbf{p}_i|| - r(s)) \, \frac{\mathbf{c}(s) - p_i}{||\mathbf{c}(s) - p_i||} - \lambda \Delta_s \mathbf{c}(s) \tag{6}$$

$$\text{and} \quad \frac{\partial r}{\partial t}(s) = -\frac{1}{N} \sum_{i=1}^{N} \Gamma_\sigma(s - \hat{s}_i) \, (||\mathbf{c}(s) - \mathbf{p}_i|| - r(s)) - \mu \frac{\partial^2}{\partial s^2} r(s) \ . \tag{7}$$

Defining a reference frame. When sweeping a 2D shape along a 3D space curve, one has to define the frame of reference, i.e. the local coordinate system for the cross-section function. This is important for building a 3D mesh and for generalizing the cross-section function.

A geometrically straightforward choice is the Frenet-Serret reference frame, as it was used in [15]. It defines the local coordinate system by the tangent, normal, and binormal vectors of the spine curve. Let again $\mathbf{c}$ denote the spine curve; we assume natural parametrization for simplicity. Then the basis vectors of the Frenet-Serret frame are defined as

$$\mathbf{Z} = \frac{\mathrm{d}\mathbf{c}}{\mathrm{d}s}, \quad \mathbf{X} = \frac{\frac{\mathrm{d}\mathbf{Z}}{\mathrm{d}s}}{||\frac{\mathrm{d}\mathbf{Z}}{\mathrm{d}s}||}, \quad \text{and} \quad \mathbf{Y} = \mathbf{Z} \times \mathbf{X}. \tag{8}$$

As mentioned above, this formulation suffers from instabilities of the second derivative, which may lead to discontinuities. We employ here a less general, but more stable reference frame. We choose a fixed vector $\mathbf{W}$ that is known to be non-collinear to the spine curve at any point. This can be done by applying prior knowledge of the organ to be modeled; in our case we chose $\mathbf{W}$ to be the normal vector of the coronal plane, in which the aorta roughly stretches. The basis vectors are then defined by

$$\mathbf{Z} = \frac{\mathrm{d}\mathbf{c}}{\mathrm{d}s}, \quad \mathbf{X} = \mathbf{Z} \times \mathbf{W}, \quad \text{and} \quad \mathbf{Y} = \mathbf{Z} \times \mathbf{X}. \tag{9}$$

3 Results

Segmentation and model fitting was tested on real time TrueFISP (SSFP) MR image series acquired orthogonal to the long axis of the aorta in three healthy volunteers, with TE 1.48, TR 2.96, FL 70, and acquisition matrix size ranging from 88×128 to 138×192; pixel size was 1.77×1.77 mm, slice thickness 6 mm. The scans were performed on a Siemens 1.5T Espree with parallel imaging (GRAPPA factor 2).

3.1 Segmentation Accuracy

The 2D segmentation was performed on a total of 2416 slices, some exhibiting low resolution, severe noise, or blood flow artifacts. Evaluation by a clinical expert showed no or very minor deviations from the true contour in 98.3% of the cases, while the remaining 1.7% were still rated clinically usable. Segmentation time per slice was less than 1 msec.

3.2 Model Fitting with Synthetic Data

To assess the accuracy of the model optimization procedure we first used a synthetically generated point cloud of known properties as data base. Uniformly distributed points were sampled on the surface of an idealized aorta model. The centerline was discretized in 35 piecewise linear segments and initialized as a straight line between start and end point. The radius function was initialized zero everywhere. Some sample images of the model after various numbers of iterations are shown in figure 3.

Model fitting was carried out for different numbers of data points, the results are shown in the table below. The detected centerline was very close to the theoretical position with a mean square error (MSE) around 10^{-5} (spine length ≈ 2.57). Denser point clouds led to a more precise result.

number of data points	100	200	500	1000
MSE $\times 10^{-5}$	4.622	3.148	2.770	2.576

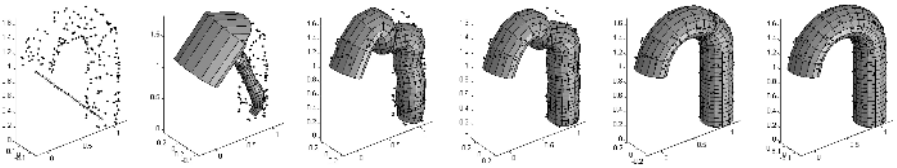


Fig. 3. Model fitting progress for synthetic data set of 200 points

3.3 Model Fitting with Real Data

Finally we evaluated the accuracy of the model on real data. Consecutive image slices were cut out from a 4D data volume and the aortic cross-section was segmented using the method described in section 2.1. The resulting model and a sample endoscopic view of it are shown in figure 4. Visual inspection showed good correspondence in most areas. Outliers did not have a big influence on the result, which is due to the smooth structure of the aorta and the model.

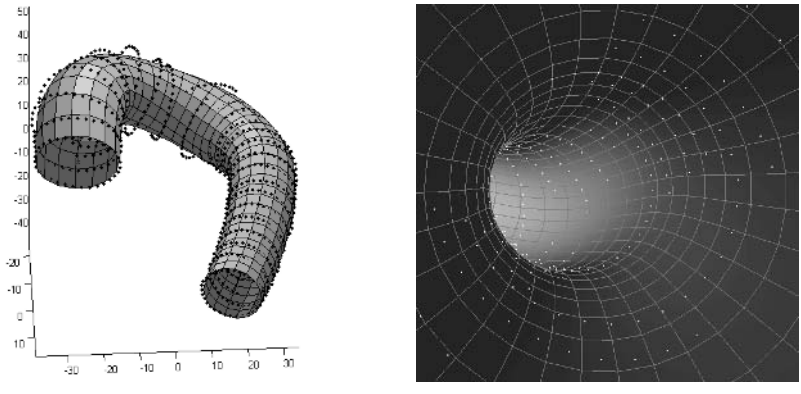


Fig. 4. Aorta model created from real human MRI data (left) and sample virtual endoscopic view (right)

4 Discussion and Future Work

We presented a method for segmenting and modeling the human aorta suitable for real-time virtual endoscopy. The highly optimized Geodesic Active Contour segmentation allows processing of an image slice in less than 1 msec. We derived and tested a variational fitting scheme for a generalized cylinder model to a dynamic point cloud. Furthermore, a new local reference frame was implemented for the model. The segmentation and model can now be integrated into the real time acquisition and feedback system described in [5].

We plan to extend the GC model by attaching a finite element mesh to the tubular GC surface to better model local deformations, which will be especially

important when applying the method to structures with more complex shape such as the left atrium.

The influence of motion was neglected in our model so far. Possible sources of displacements are the cardiac and respiratory cycles, as well as patient movement during the intervention. To cope with the periodic cardiac motion, we are extending the model to 4D with imposed constraints on timepoint-to-timepoint displacement. When addressing respiratory and external patient motion, one must consider in-plane as well as through-plane displacements. Respiratory gating methods during the acquisition as well as registration and/or quality control measures may be required depending on the exact application. And finally, we plan to develop an update strategy for the real time model based on local geometry and phase of the interventional procedure.

References

1. Zhang, Q., Wendt, M., Aschoff, A., Zheng, L., Lewin, J., Duerk, J.: Active MR guidance of interventional devices with target-navigation. Magn Reson Med **44**(1) (2000) 56–65
2. Nayak, K.S., Pauly, J.M., Nishimura, D.G., Hu, B.S.: Rapid ventricular assessment using real-time interactive multislice MRI. Magn Reson Med **45**(3) (2001) 371–375
3. Serfaty, J.M., Atalar, E., Declerck, J., Karmakar, P., Quick, H.H., Shunk, K.A., Heldman, A.W., Yang, X.: Real-time projection MR angiography: feasibility study. Radiology **217**(1) (2000) 290–295
4. Guttman, M., Lederman, R., Sorger, J., McVeigh, E.: Real-time volume rendered MRI for interventional guidance. J Cardiovasc Magn Reson **4**(4) (2002) 431–42
5. Kirchberg, K.J., Wimmer, A., Lorenz, C.H.: Real-Time Virtual Endoscopy for MR-Guided Aortic Interventions. In: Proc. 14th ISMRM. (2006) 268
6. Binford, T.: Visual perception by computer. In: IEEE Conference on Systems and Control. (1971)
7. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. International Journal of Computer Vision **22**(1) (1997) 61–79
8. Perona, P., Malik, J.: Scale-Space and Edge Detection Using Anisotropic Diffusion. IEEE Trans. Pattern Anal. Mach. Intell. **12**(7) (1990) 629–639
9. Weickert, J., Kuehne, G.: Fast methods for implicit active contour models. In Osher, S., Paragios, N., eds.: Geometric Level Set Methods in Imaging, Vision, and Graphics. Springer (2003) 43–57
10. Weickert, J., ter Haar Romeny, B., Viergever, M.: Efficient and reliable schemes for nonlinear diffusion filtering. IEEE Trans. Imag. Proc. **7**(3) (1998) 398–410
11. Adalsteinsson, D., Sethian, J.: A fast level set method for propagating interfaces. Journal of Computational Physics **118**(2) (1995) 269–277
12. Goldenberg, R., Kimmel, R., Rivlin, E., Rudzsky, M.: Fast geodesic active contours. IEEE Transactions on Image Processing **10**(10) (2001) 1467–1475
13. Krissian, K., Westin, C.: Fast and accurate redistancing for level set methods. Computer Aided Systems Theory (EUROCAST'03) (2003) 48–51
14. McInerney, T., Terzopoulos, D.: Deformable models in medical image analysis: a survey. Med Image Anal **1**(2) (1996) 91–108
15. O'Donnell, T., Boult, T., Fang, X.S., Gupta, A.: The extruded generalized cylinder: a deformable model for object recovery. In: CVPR. (1994) 174–181
16. O'Donnell, T., Gupta, A., Boult, T.E.: A new model for the recovery of cylindrical structures from medical image data. In: CVRMed-MRCAS '97. (1997) 223–232

Adaptive Script Based Animations for Intervention Planning

Konrad Muehler, Ragnar Bade, and Bernhard Preim

Department of Simulation and Graphics, Department of Computer Science,
Otto-von-Guericke University of Magdeburg, Germany
{muehler, bade, preim}@isg.cs.uni-magdeburg.de

Abstract. We describe scripting facilities to create medical animations for intervention planning based on medical volume data and derived segmentation information. A data independent scripting language has been developed to separate animation scripts from imaging data. The scripting facilities are adaptive and allow to reuse one script to create animations for many different patients. With expressive animations, we support the individual planning process, the preoperative documentation as well as discussions between medical doctors, for example in a tumor board. We also discuss the enhancement of interactive explorations with animations generated on the fly.

1 Introduction

Advances in imaging technology are continuously increasing the amount and the complexity of data and information available for intervention planning. To take advantage of these large amounts of data, the effective presentation and interaction with these data has to be considered. Currently, slice-based visualizations of medical volume data, such as CT or MRI, and derived segmentation information, are used. In very specific applications, standardized *static* three-dimensional visualizations are provided as a service [1]. Two major drawbacks of those visualizations are: (i) The difficulty to recognize exact spatial relations between the objects and (ii) the occlusion of relevant objects or parts thereof. These disadvantages can be compensated by interactive 3d-visualizations. However, navigating in 3d-scenes is time consuming and the viewer can lose sight of important objects or lose orientation [2]. Therefore, interactive exploration is rarely used in discussions between medical doctors, for example when a radiologist demonstrates a case to a surgical team.

Some of the problems of both static graphics and interactive scenes can be avoided by using carefully designed animations which effectively convey information in a limited and pre-defined period of time. The creation of expressive animations for intervention planning involves the specification of camera movements, movements of clipping planes, and the interpolation of transfer function parameters for volume rendering or parameters for surface rendering. This process is time-consuming, error-prone and requires deep and specialist skills. In intervention planning, appropriate visualizations and animations are required for

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 478–485, 2006.

many different patients. The therapeutic question, however, is often the same: for example the evaluation of resectability of a certain kind of tumor. Therefore, it is essential that animations designed for a particular case can be easily adapted or reused for another patient where the same therapeutic question is addressed. The reuse of animations is not only an issue of reducing effort but also essential to make the preoperative planning process more reproducible.

Besides the use of completely rendered animations the execution of animations during interactive exploration is conceivable. Depending on user interactions, such as selection of an object, animations can be executed which provide a scene overview or an object accentuation.

For an effective use of animations in intervention planning, we defined scripting facilities independent of a particular case, which enable an author to describe the behavior of objects in a simple and independent manner. To separate the description of animations from the individual patient data, our animation system is adaptive.

2 Related Work

The pioneering work of Zeltzer [3] introduced intent-based animation using a high-level scripting language and decomposition rules, to map higher level commands to lower level commands. This approach was used to decompose communication goals to animation instructions [4]. According to [5] animations must be slow and clear enough for viewers to perceive and understand changes. Therefore, emphasis techniques are a crucial aspect of animations [6].

For clinical purposes standardization is an important issue in the design of animations. Iserhardt-Bauer et al. [7] generated simple clinical videos efficiently using standard views and visualizations parameters. Tomandl et al. [8] evaluated the benefit of standardized high quality videos in clinical routine and state that those videos are a viable approach for standardized and reproducible visualizations. Facilities to improve interactive dataset traversal by animation of transfer functions for volume rendering were introduced by Correa et al. [9]. Krueger et al. [10], in cooperation with medical doctors, defined a standardized process for visualization of patient data in Ear-Nose-Throat-surgery (ENT-surgery). They identified default colors, texturing and viewpoints for important structures in the neck region. Preim et al. [11] introduced a scripting language for anatomy education. However, volume rendering and slice-based rendering was not considered.

Modern software libraries for scientific and medical visualizations provide facilities to create animations. In Amira [12] and Paraview [13], animations can be created for single data sets by defining events to be processed or by writing scripts in a TCL extending language. The scripts can not be applied to other data sets. Furthermore, the script instructions are very complex and difficult to handle. MeVisLab [14] provides the possibility to execute simple actions like rotations automatically. Also, a low-level scripting facility is available.

3 Script Concept

In this section, we discuss requirements for animations for clinical purposes and present a script-based concept. High-level instructions are mapped to executable low-level instructions according to decomposition rules. The selection of rules is based on an object classification (e.g. muscles, vessels, organs). Furthermore, we describe the parsing process and semantic relations to enhance animation. Finally, we present the integration of animations in the interactive exploration of a 3d scene.

3.1 Requirements

We target at *reusable* and *reproducible* animations for cases with the same therapeutic question. Thus our design is guided by four requirements:

1. **Adaptiveness:** The approximate location as well as the number of many anatomic structures, such as organs, is the same for virtually all patients. However, they differ in their specific shape and size. Thus the process of mapping high-level to low-level commands has to consider to these differences automatically. It also has to adapt to differences in pathologic structures which are unknown in size, location and shape.
2. **Convenience and specialisation:** Originating from different fields of application, the process of creating animations must strike a balance between convenience, speed and specialisation of animation.
3. **Animation techniques for intervention planning:** In addition to rotation, zoom or translation it must be possible to animate particular medical visualization techniques, such as volume rendering, clipping planes and slice-based visualizations.
4. **Standardization:** Animations for intervention planning should be standardized to afford comparisons and to facilitate the recognition of structures in different animations [7], [8], [10].

3.2 Basic Concept

We introduce a scripting language to describe animations of medical volume data and derived segmentation information. The design of our scripting language is dedicated to the requirements described above.

A script instruction consists of four parts: *time*, *object name*, *instruction name* and *parameters*. A script is executed on a visualization i.e. all instructions are applied to their associated objects over time. This can be a structure like an organ or a vessel but also the scene or the camera. The order of instructions in a script need not be the order of execution. Also an overlapping of time ranges is possible. This sample instruction moves a clipping plane through a scene from the left clipping all structures except the vessels (see also Fig. 1):

```
[0,8] 'All' clipPlane on left 0.8 except 'Vessels'
```

Fig. 1. Sample frames for clipping plane instruction above

Adaptiveness. Instructions for camera placement are relative to the corresponding object and are geared to the size and location of its bounding box. Thus the animations adapt themselves to size and location of objects. This adaption is dedicated to compact structures, such as tumors, lung nodules and lymph nodes as well as to organs. For elongated or branching structures, the bounding box is not a good approximation.

Decomposition. To afford an enhancement of the scripting language for particular therapeutical questions, we use a hierarchical approach similar to [4]. In contrast to [4] we represent (therapeutic) questions by instructions and do not abstract parts of animations like scenes or shots. By using decomposition rules, script instructions can be abstracted. We distinguish between low-level (elementary) and high-level (abstract) instructions. Low-level instructions are e.g. `move`, `rotate` and `setColor`. High-level instructions are e.g. `emphasize`, `groupOverview` and `setAsFocusObject`. By abstracting instructions, we achieve a convenient way of describing whole animations with few instructions. To support fine-grained modifications and specialized specifications high-level and low-level instructions can be mixed.

High-level commands for medical animations. In addition to instructions for object property changes and camera changes, instructions with special regard to medical visualization techniques are provided. We created an instruction to move clipping planes (see section 3.2) as well as instructions for slice-based animations and simple animations of volume rendering.

For animations of slice-based 2d visualizations, the slices containing the corresponding object are identified automatically. The `showSlices` instruction has a parameter for margin slices above and below objects' slices. The following instruction shows an animation of slices containing a tumor with 5 slices margin:

```
[0,10] 'Tumor' showSlices 5
```

According to the requirements, we also provide facilities to create animations of volume rendered scenes. We do not incorporate strongly specialized facilities to control volume rendering parameters since we do not consider to use volume rendering as the only visualization mode. Instead, we assume that volume rendering is combined with surface rendering (of segmented objects). With the

`volumeRendering` instruction, the transfer functions for gray values and transparency can be changed over time. For each transfer function, there are two parameters for *center* and *width*, to specify simple linear ramp functions. By replacing parameters with standard values they can be abstracted to parameters like *bones* or *soft tissue*:

```
[0] 'All' volumeRendering Bones
[0,5] 'All' volumeRendering softTissue
```

3.3 Script Parsing

Before its execution, a script is parsed, i.e. high-level instructions, composed objects and high-level parameters will be decomposed (Fig. 2). High-level instructions can be replaced by high-level as well as by low-level instructions. The substitution of instructions can be declared in general or depending on the anatomic type of the object the instruction refers to. Objects can be combined to composed objects so that instructions referring to a composed object are applied to all elementary objects. Parameters can be replaced by values, e.g. `red` by its rgb-value `255,0,0`. The decomposition rules are declared in XML and can be easily edited.

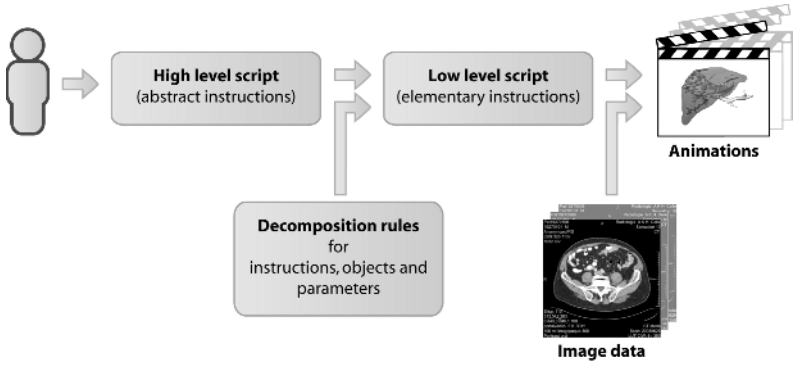


Fig. 2. Concept of transfering a high-level script to a low-level script by decomposing instructions and composed objects, and by replacing parameters. The low-level script is executed on visualization of data sets to create an animation.

3.4 Interactive Animations

Besides the use of animations as videos, animations can be combined with interactive manipulation. For the individual process of intervention planning, it is essential to provide the freedom of exploring the scene. Animations may support the exploration, e.g. by presenting an overview, an introduction of the scene or an automatic camera areal view. We consider the following scenarios for such interactive animations:

1. **Emphasis:** By selecting an object, e.g. by picking it or selecting it from a list, an animation emphasizes this object by moving the camera to a "good" viewpoint. The style of emphasis depends on the (anatomic) type of selected object.
2. **Object overview:** An overview of a selected object can animated instead of its emphasis. This can be a rotation around the object or a serial emphasis of some near focus objects.
3. **Scene overview:** To provide a re-orientation of the current position in the scene, a short animation provides a camera pan to a global viewpoint and return back to the starting position.
4. **Scene introduction:** To introduce an interactive exploration, an animation presents the scene with a rotation and an emphasis of the most important objects.
5. **Fade in and out:** If the viewer explores a selected object from different viewpoints occluding objects can be smoothly faded out when they appear in the field of view and faded in when they no longer hide the selected object.

We integrated the scripting facilities in an interactive planning system to enrich it with interactive animations.

4 Examples

Our concepts for generating medical animations are based on the analysis of numerous medical animations and fruitful discussions with medical professionals. Our scripting facilities are included in NeckVision [15], a software to segment image data and create and present visualizations for ENT-intervention-planning, and are used in the intervention planning for liver surgery. They are also a part of the LiverSurgeryTrainer [16], an educational system for liver surgeons. In the following we present two examples.

4.1 Planing a Neck Dissection

In ENT-surgery, the spatial relations in the vicinity of enlarged lymph nodes are crucial. Navigating in a scene, the surgeon can interactively explore the lymph nodes. Our animations provide amongst others a support on finding and analyzing large lymph nodes. In the example the scene is introduced and all lymph nodes are emphasized according to their size while other structures like bones and muscles are toned down to give a little context information (Fig. 3):

```
[1,7] 'Scene' introduce          // rotating the scene 360 degrees
[8,10] 'Lymphnodes' showBySize   // show lymph nodes colored by size
```

If the surgeon picks a lymph node to explore its anatomic context, this high-level script instruction will be executed:

```
[0,6] 'SELECTED OBJECT' emphasizeInContext
```

As a consequence, the structures in the near surrounding are set to nearly opaque while all other structures are set transparent. So the surgeon can analyze possible infiltrations in the muscles or vessels.

Fig. 3. Sample frames from an interactive animation of a scene introduction, an accentuation of all lymph nodes and the emphasis of a selected lymph node in its anatomic context.

4.2 Liver Tumor Resection

Before resecting a liver tumor, a surgeon has to plan where to dissect. The distance of the tumor to its surrounding vessels is an essential information in this process. With the following script, the camera is rotated around the tumor to give an overview and the camera is moved to show the minimal distance between tumor and vessels afterward (Fig. 4):

```
[0,5] 'Scene' sceneIntroduction
[6,9] 'Tumor' viewDistance 'Vessels'
```

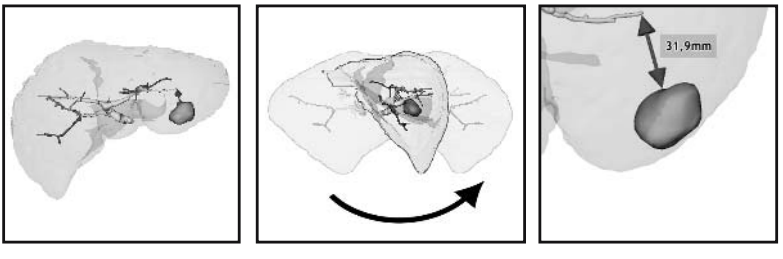


Fig. 4. Sample frames from an animation to explore a tumor and its distance to surrounding vessels. The arrow is included to clarify the rotation.

Because the scripts are reusable and adaptive the script above is used for many datasets with the medical question of minimal distance around a tumor. The animation can be played as a pre-rendered video in a tumor board discussion as well as an introduction in an individual interactive exploration process by the surgeon.

5 Conclusion and Future Work

We introduced scripting facilities which enable authors to create animations for therapeutic questions. We developed a data independent scripting language to

separate the description of animations from data sets. Because our scripting facilities are adaptive, a script can be reused for many different but similar cases. The script-based approach also allows to enhance interactive animations. The scripted animations are used to enhance important interactions with animations directly in the scene. The scripting concept can be easily extended to other therapy questions. We intend to incorporate functions to identify "good" viewpoints for objects automatically and by providing camera path planning facilities. This work was supported by the BMBF (SOMIT-FUSION project FKZ 01—BE 03B).

References

1. MeVis-Distant-Services-AG: `http://www.mevis-distant-services.com/` (2005)
2. Bade, R., Ritter, F., Preim, B.: Usability Comparison of Mouse-Based Interaction Techniques for Predictable 3d Rotation. In: Smart Graphics. (2005) 138 – 150
3. Zeltzer, D.: Task-Level Graphical Simulation: Abstraction, Representation, and Control. In: Making Them Move: Mechanics, Control, and Animation of Articulated Figures. Morgan Kaufmann, San Mateo, CA (1990) 3–33
4. Karp, P., Feiner, S.: Automated presentation planning of animation using task decomposition with heuristic reasoning. In: Graphics Interface. (1993) 118–127
5. Tversky, B., Morrison, J., Betrancourt, M.: Animation: can it facilitate? International Journal of Human Computer Studies **57** (2002) 247–262
6. Preim, B., Tietjen, C., Doerge, C.: NPR, Focussing and Emphasis in Medical Visualizations. In: Simulation und Visualisierung. (2005) 139–152
7. Iserhardt-Bauer, S., Rezk-Salama, C., et al.: Automated 3d Video Documentation for the Analysis of Medical Data. In: BVM2001. (2001) 409–413
8. Tomandl, B.F., Hastreiter, P., Iserhardt-Bauer, S., et al.: Standardized Evaluation of CT Angiography with Remote Generation of 3d Video Sequences for the Detection of Intracranial Aneurysms. RadioGraphics **23** (2003) 12e
9. Correa, C.D., Silver, D.: Dataset Traversal with Motion-Controlled Transfer Functions. In: Proceedings of IEEE Visualization 2005, Minneapolis (2005) 359–366
10. Krueger, A., Tietjen, C., et al.: Interactive Visualization for Neck Dissection Planning. In: IEEE/Eurographics Symposium on Visualization. (2005) 295–302
11. Preim, B., Ritter, A., Strothotte, T.: Illustrating Anatomic Models: A Semi-Interactive Approach. In: 4th International Conference on Visualisation in Biomedical Computing. (1996) 23–32
12. Stalling, D., Westerhoff, M., Hege, H.C.: Amira: A Highly Interactive System for Visual Data Analysis. In: The Visualization Handbook. Elsevier (2005) 749–767
13. Law, C.C., et al.: An application architecture for large data visualization. In: IEEE symposium on parallel and large-data visualization and graphics. (2001)
14. Hahn, H., Link, F., Peitgen, H.: Concepts for Rapid Application Prototyping in Medical Image Analysis and Visualization. In: SimVis2003. (2003) 283–298
15. Cordes, J., Dornheim, J., Preim, B., et al.: Preoperative Segmentation of Neck CT Datasets for the Planning of Neck Dissections. In: SPIE Medical Imaging. (2006)
16. Bade, R., Riedel, I., Schmidt, L., et al.: Combining Training and Computer-assisted Planning of Oncologic Liver Surgery. In: Bildverarbeitung fuer die Medizin. (2006)

Towards Optimization of Probe Placement for Radio-Frequency Ablation

Inga Altrogge[1], Tim Kröger[1], Tobias Preusser[1], Christof Büskens[2], Philippe L.Pereira[3], Diethard Schmidt[3], Andreas Weihusen[4], and Heinz-Otto Peitgen[1,4]

[1] CeVis – Center for Complex Systems and Visualization, University of Bremen, Germany
[2] ZETEM – Center for Industrial Mathematics, University of Bremen, Germany
[3] Dept. of Diagnostic Radiology, Eberhard Karls University, Tübingen, Germany
[4] MeVis – Center for Medical Diagnostic Systems and Visualization, Bremen, Germany

Abstract. We present a model for the optimal placement of mono- and bipolar probes in radio-frequency (RF) ablation. The model is based on a numerical computation of the probe's electric potential and of the steady state of the heat distribution during RF ablation. The optimization is performed by minimizing a temperature based objective functional under these constraining equations. The paper discusses the discretization and implementation of the approach. Finally, applications of the optimization to artificial data and a comparison to a real RF ablation are presented.

1 Introduction and Related Work

During the last decade, local and minimally invasive techniques have become a promising treatment of tumor diseases. Especially in situations where a surgical resection or a chemotherapy is not possible due to the extent of the tumor or the general state of the patient, these techniques offer an alternative and less invasive possibility of treatment. Several authors [1, 2, 3, 4, 5, 6] have worked on the simulation of various thermo-therapy approaches (intervention methods where the malignant tissue is destroyed by high or low temperatures). The complexity of the models varies from simple steady state models with constant material parameters [2] to models taking vaporization of water and the nonlinear behavior of tissue properties into account [6]. The models are discretized by finite differences or finite element methods on Cartesian or tetrahedral grids.

In this work we consider the radio-frequency (RF) ablation of liver tumors with mono- or bipolar systems: A probe, connected to an electric generator, is placed inside the malignant tissue, such that an electric current flows through the body and heats the tissue near the probe up to temperatures of more than 60 °C. At such temperatures the proteins of the heated tissue coagulate and its cells die. The treatment is successful, if the volume of destroyed tissue completely includes the malignant lesion, comparable to R0 resections in liver surgery.

Because of local cooling due to blood perfusion a thorough planning of the probe's position as well as a determination of optimal generator power are advisable to assert a successful treatment. In this paper we take a first step towards

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 486–493, 2006.

an optimization of the RF ablation planning. For a steady state partial differential equation (PDE) model we formulate an appropriate objective functional measuring the quality of an ablation. Therewith, we aim at an optimization of the RF probe's positioning under the constraining PDE system. In a future work the optimal control of the electric energy induced into the malignant tissue will be investigated.

To our knowledge, the optimization of mono- and bipolar RF ablation in three space dimensions has not been incorporated in the existing finite element models. However, for interstitial ultrasound thermo-therapy, a two-dimensional optimization model has been presented by Bustany et al. [7]. Also, a parameter identification for the hyperthermia therapy is investigated by Gänzler et al. [8]. Moreover, Villard et al [9] presented a three-dimensional optimization based on a simple estimation of the lesion zone as an ellipsoid. Our problem falls in the field of nonlinear optimization subject to infinite dimensional constraints given by a system of PDEs. For an overview of the methodology see e. g. [10, 11].

The approach presented in the current paper is the result of a joint effort of CeVis, MeVis and ZETEM at the University of Bremen and a variety of clinical partners working together in publicly funded research projects[1]. These projects focus on the development and clinical evaluation of computer aided systems for planning, monitoring, and assessing of tumor ablation. At CeVis, a group, headed by the third author, works on the advanced numerical models and algorithms.

2 A Model for the Simulation of RF Ablation

We consider the computational domain to be a cuboid $\Omega \subset \mathbb{R}^3$ with boundary $\Gamma_{\text{out}} = \partial\Omega$ in which a tumor $\Omega_{\text{tu}} \subset \Omega$ lies. Further, we assume that a mono- or bipolar RF probe is applied in Ω, whose position $\bar{x} \in \Omega$ (of the active zone's center) and direction $\bar{a} \in S^2 = \{x \in \mathbb{R}^3 : |x| = 1\}$ are the unknown optimization variables. The subset of Ω that is covered by the probe is denoted by Ω_{pr}, the subsets covered by the electrodes are denoted by Ω_+ and/or Ω_-, and $\Omega_\pm = \Omega_+ \cup \Omega_-$ (cf. Fig. 1). We emphasize that all these sets depend on $\bar{x}$ and $\bar{a}$.

In order to find the optimal values for $\bar{x}$ and $\bar{a}$, we must be able to solve the forward problem, i. e. to compute the temperature distribution resulting from the ablation for a given constellation $(\bar{x}, \bar{a})$. Hence, we are interested in the temperature T and the electric potential φ both of which are functions $\overline{\Omega} \to \mathbb{R}$. These functions are solutions of the following PDEs in Ω:

$$-\nabla \cdot \big(\sigma \nabla \varphi\big) = 0 \ , \tag{1a}$$

$$-\nabla \cdot (\lambda \nabla T) = Q \ , \qquad Q = Q_{\text{rf}} + Q_{\text{perf}} \ . \tag{1b}$$

Equation (1a) is the electrostatic potential equation derived from Maxwell's equations (cf. Stein [1]). Equation (1b) is the steady state of the heat equation,

[1] LITT / RFITT funded by the German Research Foundation (Pe 199/11-1 and Pe 199/15-2) and the national research networks VICORA and FUSION funded by the Federal Ministry of Education and Research (01EZ0401 and 01IBE03C).

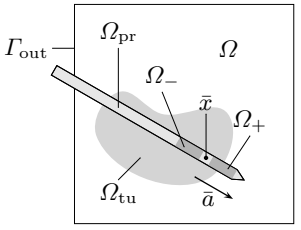


Fig. 1. Schematic setting of the considered configuration. Note that $\Omega_+ \cup \Omega_- = \Omega_\pm \subset \Omega_{\text{pr}}$, where all these sets depend on $\bar{x}$ and $\bar{a}$.

where the right hand side Q consists of the source Q_{rf} due to the electric current and the heat sink Q_{perf} due to the blood perfusion. The electric and thermal conductivities $\sigma, \lambda \in \mathbb{R}$ are assumed to be constant. Thus, we have a system of PDEs, coupled by the right hand side of the heat equation.

Both equations are considered in $\Omega \setminus \overline{\Omega_\pm}$ and Ω, respectively, together with the boundary conditions

$$\varphi = \pm 1 \qquad \text{on } \overline{\Omega_+} \text{ and } \overline{\Omega_-} \text{ (fixed potential on electrodes),} \tag{2a}$$

$$n \cdot \nabla\varphi = \frac{n \cdot (\bar{x} - x)}{|\bar{x} - x|^2}\varphi \quad \text{on } \Gamma_{\text{out}} \qquad \text{(Robin boundary conditions),} \tag{2b}$$

$$T = T_{\text{body}} \qquad \text{on } \Gamma_{\text{out}} \qquad \text{(no effect apart from the probe),} \tag{2c}$$

where n denotes the outer normal. Condition (2b) is based on the idea that on the (outer) boundary (i. e. far away from the probe), the potential behaves approximately as induced by a point load at the barycenter $\bar{x}$ of the probe.

The source term Q_{rf} is, up to a scalar multiple, equal to the electric power $P = \sigma|\nabla\varphi|^2$, i. e.

$$Q_{\text{rf}} = \frac{P_{\text{eff}}}{P_{\text{total}}} P \quad \text{with} \quad P_{\text{total}} = \int_\Omega P \, dx \ , \quad P_{\text{eff}} = \frac{4 P_{\text{setup}} R R_{\text{I}}}{(R + R_{\text{I}})^2} \ , \quad R = \frac{U^2}{P_{\text{total}}} \ .$$

R_{I} is a property of the generator (inner resistance), U is the difference of the potential φ on the two electrodes ($U = 2\,\text{V}$ for bipolar probes and $U = 1\,\text{V}$ for monopolar probes), and P_{setup} is the value set up at the generator's control unit.

Finally, the term modeling the cooling effects of the blood perfusion Q_{perf} is based on the approach of Pennes [12]:

$$Q_{\text{perf}} = -\nu \cdot (T - T_{\text{body}}) \quad \text{where} \quad \nu = \begin{cases} \nu_{\text{B}}^* \cdot \rho_{\text{m,B}} \cdot c_{\text{m,B}} \ , & \text{inside vessels} \ , \\ 0 \ , & \text{else} \ . \end{cases}$$

ν depends on the relative blood circulation rate ν_{B}^* [s^{-1}], the blood density $\rho_{\text{m,B}}$ [kg/m] and the heat capacity $c_{\text{m,B}}$ [$\text{J}/\text{kg K}$]. Actually, Q_{perf} is split up into νT_{body} (right hand side) and νT (shifted to the left hand side), thus changing the differential operator in (1b).

3 Objective Function

The aim of the therapy is the destruction of the tumor with minimum amount of affected native tissue. In this work, we focus on a temperature based objective functional, i. e. for an optimal ablation the temperature shall be high in the

region of the tumor Ω_{tu} and near body temperature in the native tissue $\Omega \setminus \Omega_{\mathrm{tu}}$. Therefore the objective functional $f \in C^1(H^1(\Omega), \mathbb{R})$ is chosen as

$$f(T) = \frac{w_1}{2} \|T - T_{\mathrm{body}}\|^2_{L^2(\Omega \setminus \Omega_{\mathrm{tu}})} + \frac{w_2}{2} \|T - T_{\mathrm{crit}}\|^2_{L^2(\Omega_{\mathrm{tu}})} \tag{3}$$

with w_1 and w_2 being suitable weights, so that the above requirements are met. Alternative approaches with objective functionals (cf. also [7]) not penalizing temperatures larger than T_{crit} inside the tumor are currently investigated. Also, coagulation based objective functionals measuring the distance between the tumor Ω_{tu} and the volume of coagulated tissue, are under consideration.

4 Optimization Algorithm

Formally, the objective function f can be considered as a function of the temperature distribution T, where T is a function of the heat source Q_{rf}, and Q_{rf} is a function of the optimization parameter $(\bar{x}, \bar{a}) =: u \in U = \Omega \times S^2$. Hence,

$$Q_{\mathrm{rf}} = \mathcal{Q}(u) \ , \quad \mathcal{Q} \in C^1(U, L^2(\Omega)) \ , \quad T = \mathcal{T}(Q_{\mathrm{rf}}) \ , \quad \mathcal{T} \in C^1(L^2(\Omega), H^1(\Omega)) \ .$$

Thus, we are looking for $u \in U$ such that $F : U \to \mathbb{R}$, $u \mapsto F(u) := f \circ \mathcal{T} \circ \mathcal{Q}(u)$ becomes minimal. Obviously, in certain situations the uniqueness of a minimizing configuration is not guaranteed, e. g. for spherical tumors. This situation may also occur in practice for liver tumors which in general have a spherical-like shape. However, such a symmetry is broken by the consideration of surrounding blood vessels and their cooling effects. Moreover, for practical reasons the uniqueness of a solution is not needed and even local minima give important information about good probe and generator configurations. In a future model we will incorporate constraints for the optimization parameters which break any existing symmetry even further. Such constraints are given by anatomical structures (bones, colon, diaphragm) that must not be punctured during the ablation.

For the optimization, i. e. the minimization of F, we use a gradient descent method. More sophisticated approaches, in particular Lagrange-Newton (SQP) methods providing superlinear convergence, are considered in an ongoing work. The current choice is due to the fact that in our case the involved Hessian turns out to be very ill-conditioned.

Thus, we construct a series $(u^n)_{n \in \mathbb{N}} \subset U$ where $u^0 \in U$ is chosen arbitrarily and $u^{n+1} = u^n + s^n v^n$. Here, $v^n \in \mathbb{R}^6$ is the descent direction which is chosen to be an approximation of $-\nabla_u F(u^n)$, and $s^n > 0$ is the step size. We determine the step size by first trying $s^n = 1$ and then bisecting it repeatedly until the resulting u^{n+1} fulfills $u^{n+1} \in U$ and $F(u^{n+1}) < F(u^n)$. The iteration continues until $\|v^n\|_{L^2(\Omega)}$ falls below a given threshold.

The remaining part of this section deals with our approximation of the six components $\partial_{u_j} F(u)$ of the gradient $\nabla_u F(u)$. Using the chain rule yields

$$\begin{aligned} \partial_{u_j} F(u) &= f'\Big(\mathcal{T} \circ \mathcal{Q}(u)\Big) \Big[\mathcal{T}'\big(\mathcal{Q}(u)\big)\big[\partial_{u_j} \mathcal{Q}(u)\big]\Big] \\ &= \mathcal{T}'\Big(\mathcal{Q}(u)\Big)^* \Big[f'\big(\mathcal{T} \circ \mathcal{Q}(u)\big)\Big] \Big[\partial_{u_j} \mathcal{Q}(u)\Big] \end{aligned}$$

where $'$ denotes the Fréchet derivative, * denotes the adjoint operator, and brackets $[\cdot]$ denote application of linear operators. For the calculation of $\partial_{u_j}\mathcal{Q}(u)$ we use a numerical approximation by central differences, whereas we determine f' analytically as

$$f'(T)[\tilde{T}] = w_1\langle T - T_{\text{body}}, \tilde{T}\rangle_{L^2(\Omega\setminus\Omega_{\text{tu}})} + w_2\langle T - T_{\text{crit}}, \tilde{T}\rangle_{L^2(\Omega_{\text{tu}})} = \langle \alpha_T, \tilde{T}\rangle_{L^2(\Omega)}$$

with $\alpha_T \in L^2(\Omega)$ given by $\alpha_T(x) = \begin{cases} w_1(T(x) - T_{\text{body}}) & \text{for } x \in \Omega \setminus \Omega_{\text{tu}} \ , \\ w_2(T(x) - T_{\text{crit}}) & \text{else} \ . \end{cases}$

For $\mathcal{T}'(\cdot)^*$ a straightforward computation leads to the representation

$$\mathcal{T}'(Q_{\text{rf}})^* \ [f'(T)] \ [\tilde{Q}_{\text{rf}}] = \langle p_T, \tilde{Q}_{\text{rf}}\rangle_\Omega$$

where $p_T \in H^1(\Omega)$ is the solution to the adjoint problem (1b), (2c):

$$-\lambda\Delta p_T + \nu p_T = \alpha_T \quad \text{in } \Omega, \qquad p_T = 0 \quad \text{on } \Gamma_{\text{out}}, \tag{4}$$

remember the term νp_T is due to the splitting of Q_{perf} introduced in Sect. 2.

5 Discretization with Finite Elements

The solutions of the elliptic boundary value problems (1) and (4) are numerically computed with a finite element method on a three dimensional uniform Cartesian grid. The electric power $P = \sigma|\nabla\varphi|^2$, which is needed to calculate the right hand side Q for (1b), is determined with central differences after the solution of (1a).

For reasons of analogy we restrict the following description to the elliptic problem (1b) which we assume to be adjusted to homogeneous boundary conditions in the usual way (see Braess [13]). We obtain the weak form, by multiplying the corresponding PDE with a test function $v \in H_0^1(\Omega)$. Integrating by parts over Ω leads to

$$\lambda \ \langle \nabla \ T, \nabla \ v\rangle_{L^2(\Omega)} + \nu \ \langle T, v\rangle_{L^2(\Omega)} = \langle Q_{\text{rf}} + \nu T_{\text{body}}, v\rangle_{L^2(\Omega)} \quad \forall \ v \in H_0^1(\Omega) \ . \tag{5}$$

In a second step we discretize this variational problem by covering the computational domain Ω with a uniform Cartesian grid. For that purpose we restrict (5) on a finite dimensional subspace $V^h \subset H_0^1(\Omega)$ consisting of piecewise trilinear, globally continuous functions. Then every function $w \in V^h$ is determined by its nodal values γ_i at the vertices x_i for $i = 1, ..., n$ of the grid. This leads to a system of linear equations, which we solve with a conjugate gradient method, implemented in a matrix free way.

6 Numerical Results

In this section we present the application of our optimization to artificial settings as well as to geometries obtained from real CT scans. Let us first verify the performance of the algorithm in a case where the correct solution is qualitatively

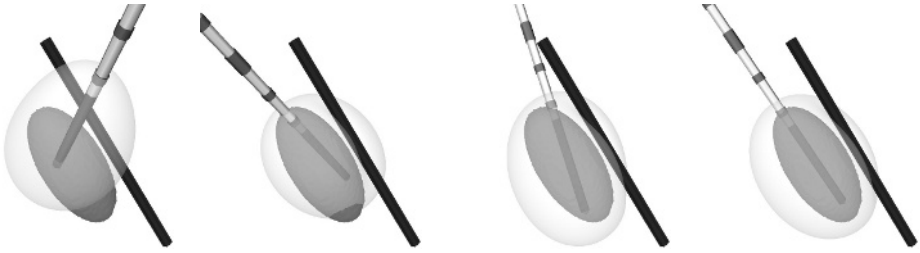


Fig. 2. Optimization of the probe's placement for an artificial example with an ellipsoid-shaped tumor (dark gray) and a linear vessel (black). The pictures show from left to right steps 0 (start), 4, 6, and 9 of the optimization and display the probe (white-gray) and a corresponding isosurface of the temperature (transparent) as well.

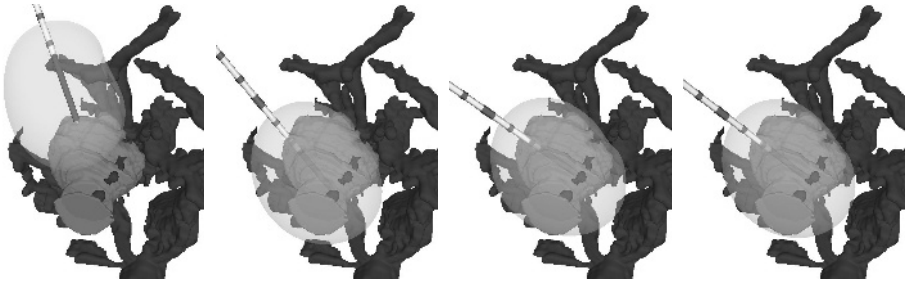


Fig. 3. Optimization for an example based on patient data with segmented tumor and surrounding vessels. The pictures show from left to right steps 0, 5, 10 and 15 of the optimization process and display the tumor together with the vascular system, the probe position and an isosurface of the temperature (cf. Fig. 2).

obvious. We let Ω be a domain of extent $60 \times 60 \times 60$ [mm^3] with an overlaid grid consisting of 121 grid points in each direction. As a tumor we consider an ellipsoid, which lies in the center of Ω, together with a straight vessel in the vicinity of the tumor (cf. Fig. 2). A monopolar probe with a radius of 1.2 mm and an electrode length of 25 mm is used. The generator has an inner resistance of $80\,\Omega$ and is set up to a power of 80 W. The initial probe position for the optimization is located at 10 mm distance in each direction from the center of Ω and its orientation is $\bar{a} = (5, 2, 3)$ normalized to length 1. The results of this experiment are shown in Fig. 2. It is clearly visible that the probe placement adapts to the ellipsoidal shape of the tumor and thus corresponds to the expected positioning along the long principal axis of the ellipsoid.

The second example is performed with a real segmented tumor and a set of surrounding vessels. The computational domain Ω has a dimension of $121.4 \times 80.7 \times 80.7$ [mm^3] and the overlaid grid consists of $155 \times 103 \times 103$ grid points. The probe data is the same as in the first example, whereas the generator settings are $R_\mathrm{I} = 80\,\Omega$ and $P_\mathrm{setup} = 30$ W. Moreover, the initial position of the probe is $(30.7, 10.3, 10.3)$mm^3 distant from the center of Ω and the probe's initial

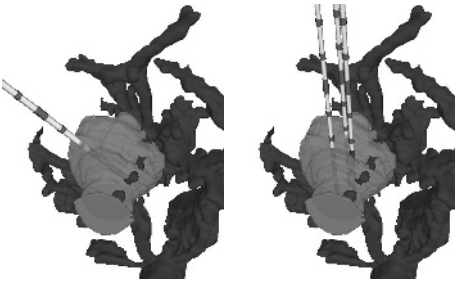


Fig. 4. Comparison of the computed optimal probe placement (left) with the probe placement chosen by the responsible physician (right)

direction is $\bar{a} = (5, 2, 3)$ normalized to 1. As the pictures in Fig. 3 show, the optimal probe's position lies in the center of the tumor, and its optimal direction adapts to the shape of the tumor as well as to the influence of the vessels. The cooling effect of the blood perfusion can be seen explicitly on the steady state of the temperature (i. e. a lower temperature near the vessels).

Finally we compare the result of our second optimization to the positioning chosen by the responsible physician. Instead of one probe of radius 1.2 mm, the physician had taken a fixed cluster of three parallel probes of radius 0.75 mm. For the time-being we assume that these act like a single probe of larger radius. Although this is only a rough approximation, the result depicted in Fig. 4 shows that the probe placement suggested by our algorithm and the one chosen by the physician nearly coincide in their position (i. e. in the barycenter of the probes' active zones), whereas the orientation differs by an angle of about 37 °. One reason for this could be the spherical-like shape of the tumor – an intrinsic property of the problem (cf. Sect. 4). Also, anatomical structures (e. g. bones or internal organs), which are not respected in our model yet, could have excluded the orientation found by our algorithm from the physician's choice.

7 Discussion and Future Work

We have discussed a model for the optimization of the placement of mono- and bipolar probes in RF ablation. The algorithm has been evaluated on an artificial example for which the optimal probe's positioning is known. Furthermore, results of the optimization with a CT-segmented tumor surrounded by a vascular system have been presented and compared to the probe's placement the responsible physician had chosen for his ablation. Although the computed optimal probe's direction differs due to e. g. anatomical structures (cf. Sect. 6), the probe's position is similar to the physician's choice.

We regard the results presented in this work as useful for the planning of RF ablations as well as for educational purposes, since they help understanding how experienced physicians choose their RF ablation parameters.

To provide practically feasible optimal positioning (respecting structures that must not be punctured), we are planning to incorporate state constraints for the optimization parameters in a future model. We regard the optimization of probe placement as a helpful tool for tumor ablation, in particular with respect to

the complete destruction of the tumor lesion. A correct placement of the probe increases the chance of a total tumor destruction and therefore reduces the risk of a recurrence. In ongoing and future investigations, we intend to incorporate sophisticated Lagrange-Newton and active set strategies for the optimization. Moreover, we will consider the corresponding time-dependent PDE models and work towards an optimal control of the overall energy deposit during ablation. In practice it will be hard for the interventionist to achieve the computed optimal positions exactly. Inevitably there will be deviations from the optimal configuration, for which a parametric sensitivity analysis is going to provide an estimation of the risk of failure of the therapy. Finally more comprehensive evaluation with artificial and real data sets (different tumor shapes and vessel configurations) will also be performed.

The authors thank the VICORA team and in particular T. Stein and A. Roggan from Celon AG for valuable hints and fruitful discussions on the topic. Also, we would like to thank the team from MeVis, especially S. Zentis and C. Hilck for preprocessing the CT scans.

References

1. Stein, T.: Untersuchungen zur Dosimetrie der hochfrequenzstrominduzierten interstitiellen Thermotherapie in bipolarer Technik. Volume 22 of Fortschritte in der Lasermedizin. Müller and Berlien (2000)
2. Tungjitkusolum, S., Staelin, S.T., Haemmerich, D., et al.: Three-dimensional finite-element analyses for radio-frequency hepatic tumor ablation. IEEE Trans. Biomed. Eng. **49** (2002) 3–9
3. Welp, C., Werner, J.: Large vessel cooling in hepatic radiofrequency ablation: investigation on the influence of blood flow rate. IEEE Trans. Biomed. Eng. (submitted)
4. Deuflhard, P., Weiser, M., Seebaß, M.: A new nonlinear elliptic multilevel FEM applied to regional hyperthermia. Comput. Vis. Sci. **3** (2000) 115–120
5. Roggan, A.: Dosimetrie thermischer Laseranwendungen in der Medizin. Volume 16 of Fortschritte in der Lasermedizin. Müller and Berlien (1997)
6. Kröger, T., Altrogge, I., Preusser, T., Peitgen, H.O., et al.: Numerical simulation of radio frequency ablation with state dependent material parameters in three space dimensions. In: Proceedings of MICCAI. (2006) submitted.
7. Khalil-Bustany, I.S., Diederich, C.J., Polak, E., Kirjner-Neto, C.: Minmax optimization-based inverse treatment planning for interstitial thermal therapy. Int. J. Hyperthermia **14** (1998) 347–366
8. Gaenzler, T., Volkwein, S., Weiser, M.: SQP methods for parametric identification problems arising in hyperthermia. Optimization Methods and Software (2004)
9. Villard, C., Soler, L., Gangi, A.: Radiofrequency ablation of hepatic tumors: simulation, planning, and contribution of virtual reality and haptics. Computer Methods in Biomechanics and Biomedical Engineering **8(4)** (2005) 215–227
10. Tröltzsch, F.: Optimale Steuerung partieller Differentialgleichungen. Vieweg (2005)
11. Geiger, C., Kanzow, C.: Theorie und Numerik restringierter Optimierungsaufgaben. Springer, Berlin (2002)
12. Pennes, H.H.: Analysis of tissue and arterial blood temperatures in a resting forearm. J. Appl. Physiol. **1** (1948) 93–122
13. Braess, D.: Finite Elemente. 3rd edn. Berlin, Springer (2003)

C-arm Tracking and Reconstruction Without an External Tracker*

Ameet Jain and Gabor Fichtinger

Department of Computer Science, Johns Hopkins University

Abstract. For quantitative C-arm fluoroscopy, we have developed a unified mathematical framework to tackle the issues of intra-operative calibration, pose estimation, correspondence and reconstruction, without the use of optical/electromagnetic trackers or precision-made fiducial fixtures. Our method uses randomly distributed unknown points in the imaging volume, either naturally present or induced by randomly sticking beads or other simple markers in the image pace. After these points are segmented, a high dimensional non-linear optimization computes all unknown parameters for calibration, C-arm pose, correspondence and reconstruction. Preliminary phantom experiments indicate an average C-arm tracking accuracy of 0.9^o and a 3D reconstruction error of 0.8 mm, with an 8^o region of convergence for both the AP and lateral axes. The method appears to be sufficiently accurate for many clinical applications, and appealing since it works without any external instrumentation and does not interfere with the workspace.

1 Introduction

C-arm fluoroscopy is ubiquitous in general surgery, interventional radiology, and brachytherapy, due to its real-time nature, versatility, and low cost. At the same time, quantitative fluoroscopy has not found a *large scale clinical acceptance*, because of inherent technical difficulties involving intra-operative calibration of model parameters, pose tracking, and target matching/reconstruction. While these aspects have been studied extensively, a clinically extant solution appears to be lacking. Advanced commercial and academic systems employ resident calibration structures [1,2,3] and optical/electromagnetic trackers or calibrated radiographic fiducials [4,5,6] to obtain the C-arm pose. The resulting equipage tends to be prohibitively expensive and complex that often interferes with the subject, image space, and clinical work-volume. While some procedures may be more tolerant to these shortcomings, despite pressing clinical needs, quantitative fluoroscopy is completely missing from brachytherapy, which is the motivating application of our project.

It can be observed that point correspondence across images (without any knowledge about their 3D locations) is a very strong constraint for pose estimation, also referred to as *bundle adjustment* in computer vision [7]. In fact,

* This work has been supported by DoD PC050170 and NIH 1R43CA099374-01.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 494–502, 2006.

six *known correspondences* across two X-ray images are sufficient to constrain the relative C-arm pose. If eight or more correspondences are available, well known linear methods exploiting the *fundamental matrix* can be applied [7], while with five correspondences, a maximum of ten degenerate solutions are possible. Thus, in general, five point correspondences across three X-ray images can recover the C-arm locations. Theoretically four correspondences have been shown to be sufficient to recover the poses in general, barring a zero measure set of the configurations that can lead to multiple solutions [8]. Though, these configurations are known to lie on certain special cubic curves (incld. cases when three points project collinearly on any X-ray image), an intuitive understanding of all these point constellations is not yet known. Moreover, it should be noted that certain special point constellations can always be created, such that they can never be resolved uniquely using any fixed number of images.

In many applications, radioactive seeds, screw/needle ends, implanted surgical markers, special anatomy points etc. are naturally present in the images. By enforcing the "consistency" of these feature points across the images, one can potentially solve for all unknown parameters of calibration, pose recovery, matching, and reconstruction in one fell swoop, in one massive high-dimensional non-linear optimization loop. In applications that do not have an adequate information in their images, one can place a few additional sticky beads or wire markers *randomly* on any temporarily stationary part (for example on the patient skin or under the operating table), and then apply the same framework. Thus the subject of this paper is a unified mathematical framework to solve the problems of intra-operative calibration, pose tracking, and target matching, and reconstruction without any sort of pre-fabricated external fiducials or tracking instrumentation. In applications where there is no need for real-time tracking of mobile surgical instruments, for example in prostate brachytherapy, the complete elimination of intra-operative tracking and calibration entourage promises to lead to a wider clinical acceptance of quantitative 3D C-arm fluoroscopy.

2 Methods and Materials

The three integral components of this problem, in decreasing order of complexity are: (1) point correspondences; (2) C-arm pose; and (3) C-arm geometry calibration. We assume that the points have been segmented from the X-ray images. Though generic and extendible to any number of images, we currently develop the framework for exactly three images. The reason for not using two images is that they have reconstruction singularities, solved by introducing a third image, which in turn makes the problem NP-Hard (*i.e.* no algorithm can even verify the optimality of a given solution in polynomial time). The theoretical complexity of four or more images is similar to that of three images. Thus we propose a detailed solution for three images, which is easily extendible to multiple images.

2.1 Mathematical Framework

Let N be the number of points chosen arbitrarily from the clinical work volume and let N_m be the number of points detected in images I_m with pose $[R_m, T_m]$ and projection model M_m. We do not assume that the 3D points are distinctly visible in all the images, but are allowed to be hidden under one another. Though this makes the correspondence problem significantly harder, it is a more realistic representation of the clinical setting. Let s_{lm} be the position of l^{th} point in m^{th} image. When three images are used, the problem can be formulated as a large optimization problem.

$$\underset{\mathcal{R},\mathcal{T},\mathcal{M},f}{\arg\min} \quad \sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\sum_{k=1}^{N_3} \mathcal{C}_{ijk} f_{ijk}$$

where

$$\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} f_{ijk} \geq 1 \;\forall\; k \;; \quad \sum_{j=1}^{N_2}\sum_{k=1}^{N_3} f_{ijk} \geq 1 \;\forall\; i; \qquad (1)$$

$$\sum_{k=1}^{N_3}\sum_{i=1}^{N_1} f_{ijk} \geq 1 \;\forall\; j \quad \textit{and} \quad f_{ijk} \in \{0,1\}$$

$\mathcal{C}_{ijk}$ is the the cost (described later) of matching point s_{i1} to points s_{j2} and s_{k3}. Note that it varies with any variation in $\mathcal{R}, \mathcal{T}, \mathcal{M}$. f_{ijk} is a discrete variable taking a value 1/0, and deciding the correctness of the match $< i, j, k >$. The inequalities force every segmented point to be chosen atleast once. Thus, f represents any feasible global match (and vice versa), with the cost of that correspondence given by $\sum\sum\sum \mathcal{C}_{ijk} f_{ijk}$. The problem hence is to compute $\mathcal{R}, \mathcal{T}, \mathcal{M}, f$ that minimize the total cost. It should be noted that since the images represent a real situation, this optimization has a solution with a near-zero cost. The only case in which a unique solution might not exist is when the information is not sufficient, *i.e.* when the number of beads are less than 7 or when they lie in a degenerate configuration.

Complexity: This is a non-linear optimization in $N^3 + 20$ variables with $3N^2$ constraints. The pose and model parameter optimization is in a continuous 20 dimensional space (2×6 for each pose, 3×3 for each model, one less for scale), while that for the correspondences is in a discrete combinatorial space. Note that we assume pixel sizes to be constant and known. Even if the pose & model parameters are known, it can be shown that the combinatorial optimization reduces to the minimum-weight tri-partite matching problem, known to be NP-Hard. This severely ill-conditions the problem, necessitating methods to constrain the problem adequately. It should be noted that though the global optima for two image matching can be proved to have only a cubic complexity, in many cases it suffers from singularities (Figure 2 (a)), forcing the use of a third image and hence an exponential complexity.

3D Reconstruction: Though crucial, 3D reconstruction is not explicitly incorporated into the framework. By optimizing for the cost, we also indirectly

compute the best reconstruction. Thus when the true pose parameters have been computed, the correct 3D reconstruction will come out as a byproduct. In the meanwhile, we shall only concern ourselves with the explicitly incorporated pose, correspondence and calibration parameters.

C-arm Imaging Model: C-arm imaging is typically approximated as a 5-parameter projection camera [1,2,3,4] to be calibrated intra-operatively for *each* individual image. Recently, however, it has been shown [9] that C-arm calibration might not always be necessary (Figure 1), implying that any reasonable calibration can be *assumed* for each image without actually calibrating at any time. Note that we assume the imager pixel sizes to be fixed and known.

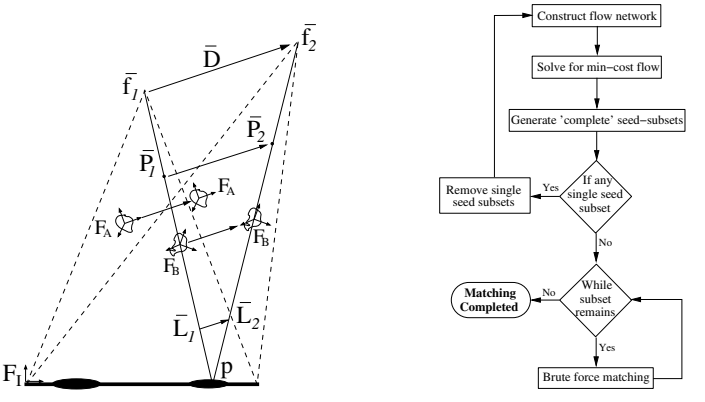


Fig. 1. Mis-calibration shifts all reconstructed objects, but keeps the relative pose nearly constant (left). The flowchart for the correspondence algorithm (right).

The central intuition is that *while an incorrect calibration gives erroneous estimates for the absolute transformations, nevertheless it still provides acceptable relative estimates.* Experimental results corroborate the theoretically derived bounds in that mis-calibration by as much as 50 mm still allows for tracking with an accuracy of 0.5 *mm* in translation and 0.5^o in rotation, and such mis-calibration does not impose any additional error on the reconstruction of small objects [9]. Thus to condition the optimization in Equation (1) better, it it advisable not to solve for the imaging parameters using sparse data from the image, but just use nominal values that may be known from a pre-calibration or the manual/header. An alternate perspective is to notice that since reconstruction errors change only negligibly with calibration errors, any attempt to calibrate using a sparse point set (instead of a very accurate calibration fixture) will ill-condition the problem by allowing for a whole space of feasible solutions. Thus it is wiser to fix the value at a choice that is practically close to reality. Later, if needed, the calibration can be further refined after the optimization converges.

Though we do not explicitly address the issue of distortion correction, advancements in intensifier tubes allow for lesser distortion and more recently the

advent of flat panel detectors obviates this step altogether. Furthermore, many application like brachytherapy, with their limited C-arm workspace, allow for a constant pre-operative distortion correction.

Correspondence: Assuming known pose parameters, we converted the point correspondence problem to a weighted tri-partite matching problem in Equation (1), an NP-Hard combinatorial optimization problem. An attractive approximate solution using a network-flow-based combinatorial optimization has recently been extended to efficiently deal with "hidden seeds", (i.e. points that overlap in some images) in practically $O(N^3)$ times [10].

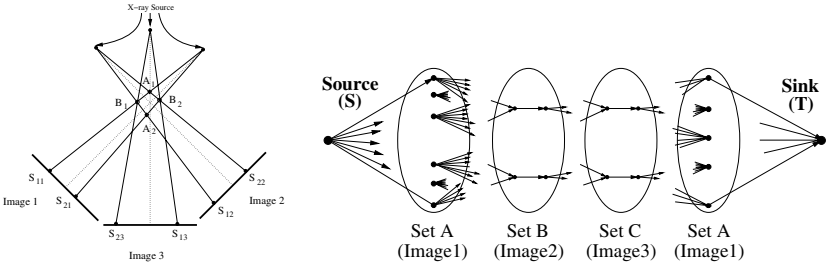


Fig. 2. A third image is needed to resolve two-image singularities (left). The flow network formulation used to solve the correspondence problem.

Sets A, B, C, and D, represent an image each. Links with a cost connect every feasible match between any two images. A flow of value N originates at the source S and ends at sink T. The problem reduces to computing a min-cost flow, easily computed using the cycle cancelling algorithm (pushes negative cycles until there are none left). A flowchart is illustrated in Figure 1. Since the problem is NP-hard the network cannot completely constrain the same point in both set A and D. Nevertheless, it works well, producing near perfect matchings.

Cost Metric: In general, any cost-metric that directly measures the deviation from the observation should perform well. The metric should incorporate all the available information, making the global minimum sharper and the algorithm robust. One good choice for a metric is *projection error* (PE). For any given set of poses and correspondence, the intersection of the three lines that join each projection to its respective x-ray source can be computed using a closed form solution that minimizes the L_2 norm of the error. PE can be computed by projecting this 3D point in each image and then measuring the distance between the projected location and the observed location of the point.

2.2 Optimization Strategy

Due to the convolution of both continuous and discrete parameters, the optimization in Equation (1) becomes ill-conditioned. Incorrect pose estimates will

invoke incorrect correspondences, especially for *dense* point clouds. However, it should be observed that *incorrect pose estimates and the subsequent correspondences invoked by it, are typically inconsistent*. This is because, while it is true that any given pose estimates invoke a correspondence, it is also true that any given correspondence also invokes a unique pose where Equation (1) will be minimized for that correspondence. Thus the desired minima will be such that the current pose invokes a correspondence, and the correspondence in turn will also invoke the same pose (with near zero cost). This order of stability we believe, will exist only at the true unique global minima. We propose an iterative strategy that exploits this observation, in spirit similar to a coordinate descent method.

Another observation to make is that given any generic estimate of the pose, the correspondence is usually completely incorrect. Nevertheless, some other pose in the vicinity can usually establish atleast a few correct correspondences ($\sim 10\%$). This new pose estimate will *behave* like a local minima. We say behave because, for any fixed pose estimate, computing f such that Equation (1) is minimized is not a polynomial time computation ($N!^2$). Thus only a working algorithm can be practically available. Even though our correspondence algorithm has been experimentally shown to be over 98% accurate near the correct pose, its assumptions start breaking down at incorrect estimates. Nevertheless, if we can estimate these few correctly matched points and block the rest, we can quickly converge to the correct answer. In the absence of any additional information, a working strategy is that if a flow in the network originates at a vertex i in set A and also ends up at vertex i in set D, then this flow is *self-consistent*. We choose a subset of self-consistent points as matched points (typically the ones closer to the average PE). These few points can now be easily used to update the pose estimates, which in turn could provide a improvement in the correspondences at a later stage. Thus the algorithm iteratively establishes the best possible correspondences (keeping the pose relatively constant) and then uses a *self-consistent* subset of points to refine the pose. As the iteration progresses, it can stop only at a self-consistent parameter choice where the pose and correspondence perfectly complement each other.

For very high density clouds, the optimization might require non-practical number of iterations to converge. Two main methods to constraint such cases are: (1) establishing a good initial estimate using prior knowledge about the surgical protocol and workspace constraints. (2) in the absence of good initial estimates, a couple of known correspondences can prove to be sufficient. These might be naturally available or artificially induced. In any case, this might become a necessary step since projective geometry can recover the 3D reconstruction only up to an arbitrary scale. To recover the scale, information external to the image is required (ex. length of an inserted screw), allowing for a few known correspondences.

3 Phantom Experiments and Results

A radiographic fiducial was used to track the C-arm (0.56 mm translation; 0.33^o rotation accuracy), and was accurately attached to a point cloud phantom as

shown in Figure 3. The cloud phantom comprises of multiple slabs, thus capable of multiple *random* point configurations. 100 points with 1.56 *points/cc* were used. X-ray images within a 20^o cone around the AP-axis were *randomly* taken using an *Philips Integris V3000* fluoroscope and dewarped. Thus both the seed locations and X-ray pose were not biased/optimized in any way, closely representing an uncontrolled surgical scenario. Each image was hand segmented to establish the true segmentation and correspondence. The true C-arm pose and reconstruction was compared to that computed from the algorithm.

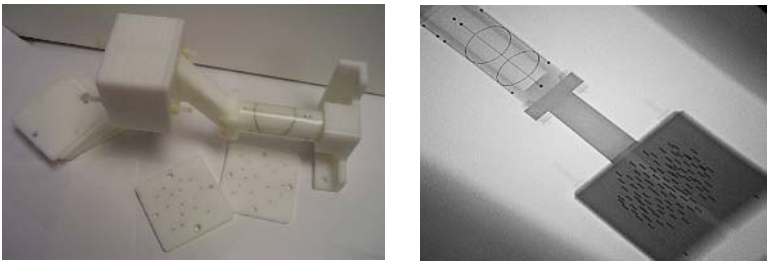


Fig. 3. An image of the point phantom attached to the fiducial (left). The phantom can replicate many point configurations. A typical X-ray image of the combination (right).

We divided the experiments into three separate cloud sizes: sparse having an average 3D point separation $\geq 25\ mm$, medium having $\sim 15\ mm$, while a high density one being $\leq 10\ mm$. These represent different types of surgical scenarios, ranging from orthopedic to brachytherapy. We generated random clouds using $10-20, 20-40$ & $40-100$ points. Region of convergence (ROC), accuracy of C-arm tracking and reconstruction error (RE) are the three metrics used to evaluate performance. Since the scale is not directly recoverable, only the rotation errors are used to study errors in the pose. To study RE, the scale is established using two points from the fiducial.

Figure 4 plots the performance of the algorithm. Each data point is averaged using 10 random runs of the initial estimate and the point cloud. When no prior correspondences are available, the algorithm could have some difficulty in converging reliably. Pose recovery accuracies vary with the number of available points, the average being about 0.9^o, while RE remains fairly stable at $0.8\ mm$. The ROC for sparse and medium sized clouds is $8-10^o$ (individually along both AP and lateral axes), while it is about 6^o for dense implants.

Runtime: The algorithm was implemented in Matlab 7 on a Windows PC (3.2 GHz P4, 1GB RAM). The algorithm would typically converge in anywhere between 3-7 total iterations, taking 2-7 minutes, depending on the point cloud density and initial estimate. However, it should be noted that it spends about 50-70% of the time for file I/O (a Matlab constraint). Thus a C/C++ implementation is expected to run in 30s.

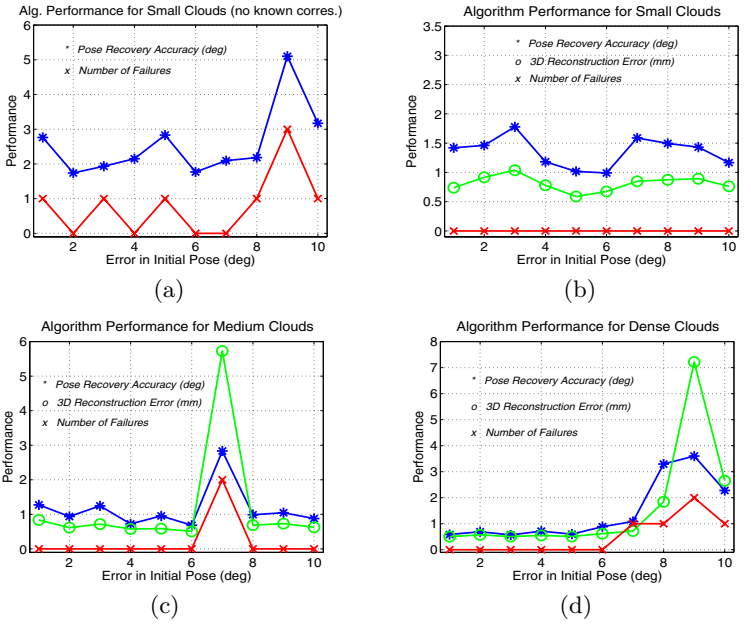


Fig. 4. The performance of the algorithm as a function of initial estimate. (a) is with no known correspondences, while (b)-(d) are with a few known ones.

4 Conclusion

A unified framework for point correspondence, C-arm tracking and reconstruction has been proposed and experimentally validated on phantoms. The experiments indicate an accuracy of 0.9^o for tracking, 0.8 *mm* for 3D reconstruction and a convergence region of 8^o (each) in both the AP and lateral axis of rotation. The framework does not need external fiducials for C-arm pose estimation and is capable of using information naturally present in the X-ray images of a family of clinical applications, such as prostate brachytherapy. In applications where this information is not present, or a greater accuracy is desired, the framework easily extends by randomly attaching beads around the patient. Our technique does not compromise on the available clinical work volume. The framework can also accommodate any available prior information on projection angles or correspondences to constrain the optimization better, and thereby to achieve a higher accuracy.

The main concern for the clinical use of methods relying heavily on high dimensional optimization is that of providing uniformity and reliability in performance. We have conducted our validation on randomly selected views and number/distribution of the points, indicating the robustness of the algorithm to these issues. Nevertheless, this is only a first step and further work to achieve better uniformity in the results is desirable. Though the alternative of well designed

calibration fixtures and image acquisition procedures are also available, they become cumbersome in many procedures. Further development of the current approach or even an amalgamation of the two approaches, could significantly improve the current clinical viability of intra-operative quantitative fluoroscopy. The development of a clinical prostate brachytherapy system to further validate our approach is currently underway. Note that, even though the driving application was prostate brachytherapy, the method also has potential in many synergistic applications in orthopedics and angiography.

References

1. Navab, N., et al: Dynamic geometrical calibration for 3-d cerebral angiography. In: SPIE Medical Imaging. (1996) 361 – 70
2. Tang, T.: Calibration and point-based registration of fluoroscope images. Master's thesis, Queen's University (1999)
3. Livyatan, H., Yaniv, Z., Joskowicz, L.: Robust automatic c-arm calibration for fluoroscopy-based navigation: A practical approach. In: MICCAI. (2002) 60–68
4. Hofstetter, R., Slomczykowski, M., Sati, M., Nolte, L.: Fluoroscopy as an imaging means for computer-assisted surgical navigation. CAS **4(2)** (1999) 65–76
5. OEC 9800 FluoroTrak^{TM}: (GE Healthcare, Waukesha, WI)
6. Jain, A., Mustufa, T., Zhou, Y., Burdette, E.C., Chirikjian, G., Fichtinger, G.: A robust fluoroscope tracking (ftrac) fiducial. Med Phys **32** (2005) 3185–98
7. Ma, Y., et al: An Invitation to 3-D Vision. Springer (2000)
8. Holt, R and et al: Uniqueness of solutions to three perspective views of four points. IEEE Trans. on PAMI **17** (1995) 303–307
9. Jain, A., Kyon, R., Zhou, Y., Fichtinger, G.: C-arm calibration - is it really necessary? In: MICCAI. (2005) LNCS 3749, 639–646
10. Kon, R., Jain, A., Fichtinger, G.: Hidden seed reconstruction from c-arm images in brachytherapy. In: IEEE ISBI. (2006) 526–29

Rigid-Flexible Outer Sheath Model Using Slider Linkage Locking Mechanism and Air Pressure for Endoscopic Surgery

Akihiko Yagi[1], Kiyoshi Matsumiya[1], Ken Masamune[1], Hongen Liao[2], and Takeyoshi Dohi[1]

[1] Graduate School of Information Science and Techonology, The University of Tokyo, 7-3-1 Hongo Bunkyo-ku, Tokyo, 133-8656, Japan
{akihiko, mkiyo, masa}@atre.t.u-tokyo.ac.jp, takdohi@i.u-tokyo.ac.jp
[2] Graduate School of Techonology, The University of Tokyo, 7-3-1 Hongo Bunkyo-ku, Tokyo, 133-8656, Japan
liao@miki.pe.u-tokyo.ac.jp

Abstract. The objective of this paper is to develop an outer sheath for flexible endoscopic manipulators. This sheath can switch two states including flexible and rigid, and make a rigid curved path for inserting manipulators. The flexible mode can be curved into a required shape. The rigid mode can hold the shape of the sheath, and then keep the path for instruments. Through the managed path, the flexible manipulators become easy to reach the target. We proposed a serial multi joint model to realize the flexible mechanism. This model is composed of a set of frame units which are connected serially. Each unit can be rotated to a given angle around the center of the joint. We developed a slider-link mechanism and a gear stopper controlled by air pressure for rigid mode. We designed and fabricated the prototype with a diameter of 16mm and length of 290mm. The experiment showed that the device could be switched from the flexible mode to the rigid mode when the air pressure was over 150kPa, and each joint could hold its angle against the maximum 400mNm. The phantom experiment showed that the flexible devices are possible to transmit the wire tension to the endpoint of the manipulator without changing the curving shape with by the developed outer sheath device.

1 Introduction

Laparoscopic surgery has been widely performed in abdominal surgery. Surgeon can insert special instruments from small holes on the abdominal wall and perform operation after lifting the abdominal wall. To improve the performance and reduce the risks of this surgery, many researchers and engineers developed various instruments such as multi degrees-of-freedom (DOFs) bending manipulators [1]-[3] or Master-Slave type manipulators[4]. However, some problems remain unsolved. First, laparoscopic surgery requires wide space below the abdominal

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 503–510, 2006.

wall. Pneumoperitoneum is commonly used to lift the abdominal wall, but some complications caused by this method are reported. Second, although the laparoscopic surgery can approach the target in anterior area of the body, it is difficult to approach the target in deep, narrow area. To address these problems, instruments are required to be flexible to go into the body avoiding the critical areas. Several groups developed different flexible manipulators with wide curve. Ikuta et al. developed a micromanipulator to reach difficult area[5]. Moreover, other flexible manipulators using shape memory alloy (SMA)[6], or wire-driven mechanism were also developed[7]. However, flexible instruments can not be inserted easily in narrow space between the tissues or organs, and can not be stabilized completely when approaching target. In order to solve problem of flexible instruments, it is required to manage rigid path to insert flexible instruments in advance.

The purpose of this study is to develop an outer sheath to make an instrumental path in human body. This sheath is able to be any given shape, and to hold the shape against of the external force. Before inserting flexible instruments, the surgeon inserts the outer sheath through the narrow gap between the safety areas. After reaching to the target, the surgeon locks the shape, then the surgeon is able to insert flexible instruments easily through the planned path. In order to realize the switch of two states, flexible and rigid, we used multi-joint model for flexibility and slider-linkage mechanism and air pressure locking mechanism for rigidity. The engineering contribution of this paper is to devise a mechanism which can change its shape to be given shape and generate stiffness in the given shape. The clinically significance is to enable to manage rigid curved path for flexible instruments. Then, it becomes possible to approach to the target by less invasive way. This paper reports 1) a mechanism that enables to switch flexible mode and rigid mode, 2) a prototype of outer sheath and total system, and 3) evaluation experimental results about the stiffness and clinical applicability study using phantom and flexible device.

2 Method

The outer sheath that we try to develop is able to change its shape with a number of DOFs. To realize this function, the multi joint model is used[8]. The outer sheath consists of several pieces of short sheath. These sheath units are connected serially to make a long sheath. Each unit is able to be rotated several degrees around the joint. So this kind of outer sheath is possible to be changed to different (Fig. 1). To lock the shape and keep from the external force, we proposed slide-linkage locking mechanism. Each unit has a link, a slider, a stopper, and an air channel inside the instrument. The unit and the slider are connected by a link with pin joints. The slider moves in the slider channel of the adjacent unit in conjunction with the rotation of the unit. Sliders and the stoppers have gear cogs. When air pressure pushes the stopper, the stopper meshes the gear tooth of the slider. The rotation of the unit is locked after locking the slider, so the sheath gets rigid. When the air pressure is off, the spring attached to the stopper

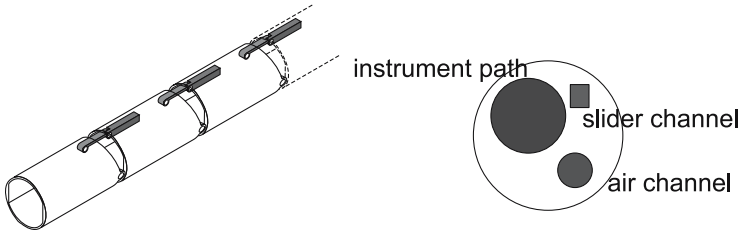


Fig. 1. The perspective view (left) and cross section image (right) of the outer sheath model: Our outer sheath consists of pieces of the sheath, link, and slider. Link and slider slide in slider channel and flexible tube passes through air channel. Instruments pass through the instrument path.

push the stopper back, and then the sheath is switched to the flexible mode (See Figure 2).

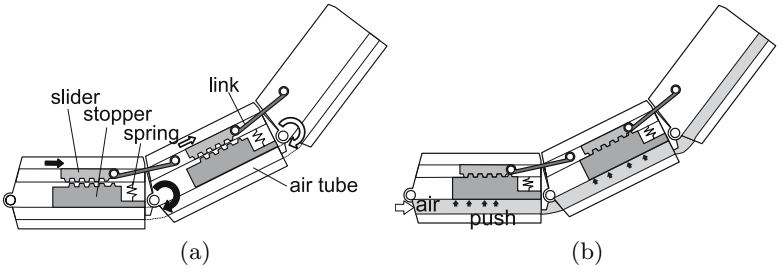


Fig. 2. Mechanism of having stiffness:When fluid channel is empty, the piece, link, and slider move in conjunction, and each pair moves independently (a). When the air pressure is add, the stopper moves up and mesh with the gear rack of the slider, then the shape of the sheath is locked (b).

3 ProtoType

We made a prototype using above mechanism(Fig. 3). Long units and short units are connected alternately, and two joints of each unit have different direction of rotation axis. Then the sheath is able to make cubic curve. The cross section of this sheath is circle diameter of 16 mm. Diameter of the instrument path is 8mm. Polyvinyl chloride tube is attached in the inner guide tube of this hole, so the sheath enables instruments to go through the device with a diameter of less than 6 mm. Diameter of the air channel is 3.6 mm. The natural rubber tube is attached to transmit air pressure to the stopper. The length of long unit is 22mm, and that of short unit is 7mm. Total length is 290 mm. Constituent material of the unit frames is aluminium. The links, sliders, and joint pins is made of stainless-steel. The total weight is 98g.

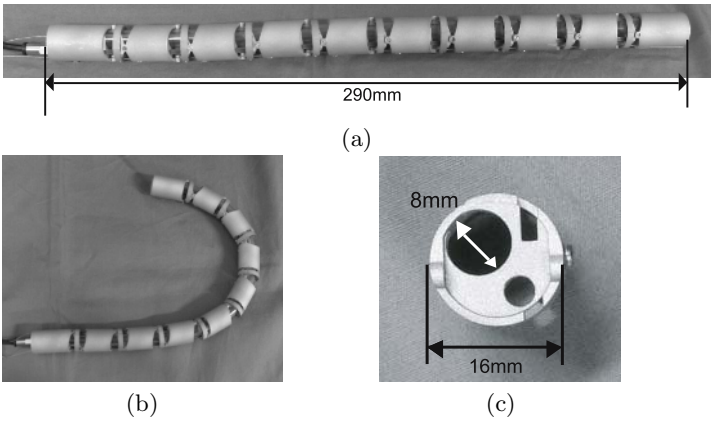


Fig. 3. The image of prototype: (a) shows when this outer sheath becomes straight. (b) shows that this out sheath is possible to curve and lock the shape. (c) shows the cross section surface.

The prototype has one pair of wire guides periodically spaced along its length. Two wires run through the wire guide eyelets, terminating at the endpoint. Through the wire tension, the sheath curves and change direction of the endpoint such as planar continuum manipulators[9] when the sheath is flexible. The wire is pulled by hand, and this wire tension supports to going into the sharp curved route and selects the planned route from branched route.

Figure 4 shows the system configuration. This system consists of the outer sheath device and air pressure controller. The controller switches to supply air pressure and controls the pressure. Air source is supplied from air supply port in operating room, gas cartridge, or simple air pump. This prototype the hand type switch is used, but any type, such as foot switch, is usable. This sheath is used under draping. In the future the wire tension form motors and motor drivers, then the outer sheath will be controlled from PC.

4 Validation

We evaluated the basic mechanical characteristic and clinical applicability. First we evaluated the relationship between the air pressure and the ability to keep the shape against external force (external torque). External force is added to the head of the sheath perpendicular to the direction of first unit (In this paper we call the pieces and joints sequential order form here.). Then, we measured the external force that rotates the first unit around the first joint, with a air pressure change from 0kPa to 400kPa. Figure 5 shows two clear states switching on the certain border air pressure. When the air pressure is low, the external force is about 0gf, this means the first piece rotated freely. When the pressure is high, the unit holds its angle even the external force is add. The border zone of the air pressure is from 130kPa to 150kPa.

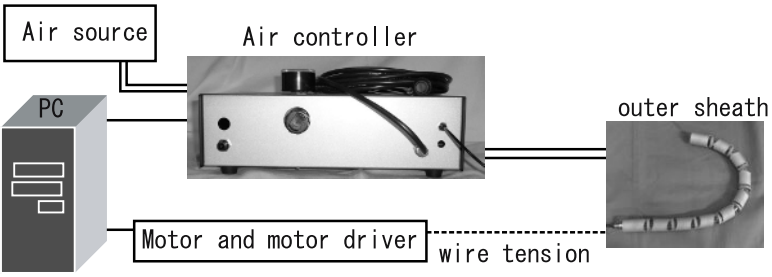


Fig. 4. The system of rigid-flexible outer sheath: The system consists of the outer sheath and air controller. Single lines are the connection of electric signals. Double lines are connection of air. Dashed line is the image of wire. Future, the wire tension will be controlled from motor and PC.

Next, we measured the external force by keeping the angle of each joint in the same condition about external force. The air pressure was set to 250kPa. In this experiment the shape of the sheath was straight. The first joint and second joint keep their angle against over 5N, but third joint bends when the external force is 4.5N. Maximum force becomes lower when the distance between the joint and the forced point is long. As the result, the maximum torque should be 400mNm.

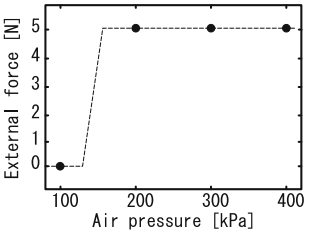


Fig. 5. Relation between rigidness and air pressure: Vertical axis is the external force under which first joint could hold its locked angle. The force was added perpendicular to the direction of the axis of the sheath. We tested the external force when the air pressure is added every 100 kPa, and every 10kPa between 100kPa and 200kPa. When the air pressure is between 130kPa and 150kPa, the rigidness is unknown.

Last, we evaluated clinical applicability by the phantom experiment. As the alternative of human soft tissues and gap between organs, narrow curved path was made and the silicone gel was filled around the path(Fig. 6a). In this experiment the path was the arc of which curvature radius is 10cm. When the outer sheath is inserted, it has to push the silicone wall to go through the path and reach to the target. In this experiment we assume that the sheath and manipulators go to the backside of the liver passing around liver and cholecyst (Fig. 6c), so the silicone is shaped like liver. As a flexible device, we used the laser

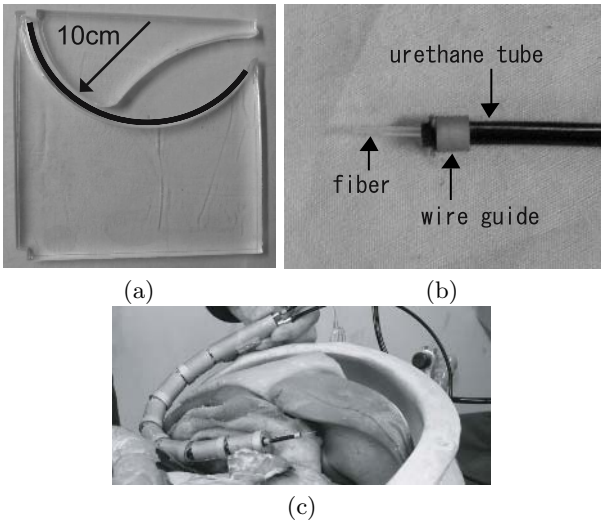
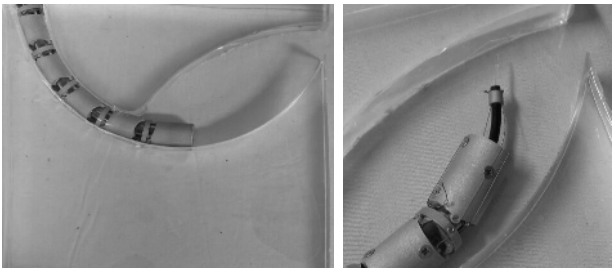


Fig. 6. Experiment condition: (a) curved course made by gelatin. The curveture radius is 10cm, and the thickness is 25mm. (b) insert material. The diameter of fiber core is 1mm. Urethane tube is 3mm, and that of wire guide is 4.5mm. (c) assumption of the phantom. This phantom assumes that the sheath goes through the gap around liver and cholecyst, and reaches to target on the backside liver.

fiber. The fiber(Standard-Lightguide, Dornier MedTech inc, German) is covered with urethane tube and wire guide device is attached periodically spaced. Two wires runs though with the eyelet of wire guide attachments. This fiber can be controlled by the difference of wire tension. The maximum diameter of the fiber device is 4.5mm(Fig. 6b). The outer sheath wasn't able to go through the path only by pushing the sheath because the top of sheath was blocked by the silicone wall. However, with supporting the wire tension, it could go the path smoothly the path. (Fig. 7). After inserting the outer sheath device, the laser fiber device was inserted smoothly passing through the rigid path. When wire tension was add, the fiber device changed its direction of the tip without changing the whole shape of the device, so this manipulator could be controlled the direction of laser. The wire tension to bend the fiber was 18N.

5 Discussion

Experimental results show that slider linkage locking mechanism is sufficient to fix the shape. Laparoscopic manipulators can generate about 500gf (4.8N) force perpendicular to the direction of axis to lift up heavy organs in the body. The outer sheath keeps its shape against such force. Since our manipulators can go through the narrow gap between tissues, it is not necessary for the surgeon to move the organs. As for the air pressure, the outer sheath switches the pressure less than 200kPa. It is not high pressure, so the air pressure may not damage



(a) (b)

Fig. 7. The image of experiment: (a) shows that the outer sheath went through the path pushing silicone walls. (b) shows the endpoint of the sheath and laser fiber curved by wire tension.

the tissues around the sheath if the air leak, or burst. Then it can be used safely. From the phantom experiment, this sheath is considered to be able to go back side passing around the tissues. The arc whose radius is 10cm is a little large path to go around small organs. The sheath isn't able to go large curvature route, or bend the head like bending manipulators. Then it is difficult to go around the small organs. However, it is sufficient to go between abdominal wall and tissues, or go thorugh the subxiphoid route. Wire tension is helpful to go the sharply curved route. In this prototype the tension affects to all joints, but the wire tension is helpful for only the first joint. By the mechanism to transmit the wire tension to first joint, the sheath will be easier to go narrow route. Using this sheath, the manipulators can be inserted smoothly, and transmit the wire tension to the endeffector. In phantom experiment we used a laser manipulator. We think that the gastroenterological endoscope is also useful combined with the sheath.

Next challenging problem is to detect the place of the endpoint of the sheath in human body, especially while inserting this sheath. X-ray is one effective way, but exposure problem occurs. Then smart way to detect the position is required.

We think that it become to be able the endoscopic surgery in deep, narrow area by the combination of our outer sheath and the wire-driven flexible manipulators. Our research is important step to make surgery less invasive.

6 Conclusion

We developed a rigid-flexible outer sheath that secures instrumental path for flexible device to make minimally invasive laparoscopic surgery. The mechanism to switch flexible mode and rigid mode using multi-joint model and slider-linkage mechanism is developed and evaluated. The air pressure and gear lock mechanisms enable a freedom shape locking. The results show that the outer sheath is able to bear about 5N force on head, and 400mNm torque. Furthermore, flexible devices are possible to change the direction of its endpoint without changing

the whole curving shape. Our mechanism that combined of the outer sheath and flexible manipulators is suitable to the laparoscopic surgery in deep, narrow area where the open surgery is necessary.

Acknowledgment

A part of this work is supported by Research Program for development of analysis, help, and alternation of human function "Research for the development of new surgical robot" , administrated by Ministry of Health, Labour and Welfare of Japan.

References

1. Peirs, J., Reynaerts, D., Van Brussel, H.: A miniature manipulator for integration in a self-propelling endoscope. Sensors and Actuators A. **92** (2001) 343–349
2. Nakamura, R., Oura, T., Kobayashi, E., Sakuma, I., Dohi, T., Yahagi, N., Tsuji, T., Hashimoto, D., Shimada, M., Hashizume, M.: Multi-DOF Forceps Manipulator System for Laparoscopic Surgery - Mechanism miniaturized & Evaluation of New Interface -. Proc. of 4th International Conference on Medical Image Computing and Computer-Assisted Intervention. (2001) 606–613
3. Yamashita, H., Hata, N., Kim, D., Hashizume, M., Dohi, T.: Handheld Laparoscopic Forceps Manipulator Using Multi-slider Linkage Mechanisms. Proc. of 7th International Conference on Medical Image Computing and Computer-Assisted Intervention. **2** (2004) 121–128
4. Abbou, C.C., Hoznek, A., Salomon, L., Olsson, L.E. Lobontiu, A., Saint, F., Cicco, A., Antiphon, P., Chopin, D.: Laparoscopic radical prostatectomy with a remote controlled robot. JOURNAL OF UROLOGY. **165** (2001) 1964–1966
5. Ikuta, K., Sasaki, K., Yamamoto, K., Shimada, T.: Remote Microsurgery System for Deep and Narrow Space - Development of New Surgical Procedure and Microrobotic Tool. Proc. of 5th International Conference on Medical Image Computing and Computer-Assisted Intervention. (2002) 163–172
6. Nakamura, Y., Matsui, A., Saito, T.: Shape Memory-Alloy Active Forceps for Laparoscpic Surgery. Proc. of the 1995 IEEE International Conference on Robot and Automation. (1995) 2320–2327
7. Simaan, N., Taylor, R., Flint, P.: High Dexterity Snake-Like Robotic Slaves for Minimally Invasive Telesurgery of the Upper Airway. Proc. of 7th International Conference on Medical Image Computing and Computer-Assisted Intervention. **2** (2004) 17–24
8. Okamoto, J., Iida, M., Nambu, K., Fujie, M., Umezu, M.: Development of Multi-DOF Brain Retract Manipulator with Safety Method. Proc. of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems. **3** (2003) 2594–2599
9. Ian, A.D., Christophoer, D.R., Ian, D.W.: Large Deflection Dynamics and Control for Planar Continuum Robots. IEEE/ASME Transactions on Mechatoronics. **8(2)** (2003) 299–307

Combined Endo- and Exoscopic Semi-robotic Manipulator System for Image Guided Operations

Stefanos Serefoglou[1], Wolfgang Lauer[1], Axel Perneczky[2], Theodor Lutze[3], and Klaus Radermacher[1]

[1] Helmholtz Institute for Biomedical Engineering of the RWTH Aachen University, Pauwelsstrasse 20, 52074 Aachen, Germany
{serefoglou, lauer, radermacher}@hia.rwth-aachen.de
http://www.meditec.hia.rwth-aachen.de/de/index.shtml
[2] Neurosurgical Department of the Johannes Gutenberg-University of Mainz, Langenbeckstrasse 1, 55101 Mainz, Germany
per@nc.klinik.uni-mainz.de
http://www-klinik.uni-mainz.de/Neurochir/index.html
[3] Aesculap AG & CO. KG, Am Aesculap Platz, 78532 Tuttlingen, Germany
theodor.lutze@aesculap.de
http://www.aesculap.de

Abstract. This paper describes the development of a robotic assistance system for image guided operations. To minimize operation time, a multimodal user interface enables freehand robotic manipulation of an extracorporeal stereoscopic digital camera (exoscope) and an endoscope. The surgeon thereby wears a head-mounted unit with a binocular display, a head tracker, a microphone and earphones. Different view positioning and adjustment modes can be selected by voice and controlled by head rotation while pressing a miniature confirmation button with a finger. Initial studies focused on the evaluation and optimization of the intuitiveness, comfort and precision of different modes of operation, including a user test with neurosurgeons in a virtual reality simulation. The first labtype of the system was then implemented and demonstrated in the operating room on a phantom together with the clinical partners.

1 Introduction

Minimally invasive surgery is typically performed through several small incisions using slender instruments and a rigid endoscope. Compared to traditional open surgery, it reduces patient trauma and recovery time but also the surgeon's dexterity and sensibility. Visual assistance systems are necessary whenever the direct sight of the surgeon to the surgical field is difficult or impossible, and are useful to augment images with e.g. preoperatively segmented outlines or navigation data. Several systems utilize robotic manipulators and have already been proposed or used in surgery, as listed in [1, 2] and Table 1. However, some of these, especially retrofitted industrial robots, obstruct the workflow through large size and/or cumbersome handling, thus reducing their acceptance by the clinical staff.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 511 – 518, 2006.

Table 1. Representative robotic manipulators for intraoperative visual assistance

System	Institute/Company	Manipulation	Ref.
AESOP (ZEUS)	Computer Motion	endoscope (& instruments)	[3]
CLEM	TIMC/IMAG	endoscope	[4]
da Vinci	Intuitive Surgical	endoscope & instruments	[5]
EndoAssist	Armstrong Healthcare	endoscope	[6]
FIPS Endoarm	MIC, Tübingen Univ.	endoscope	[7]
KaLAR	KAIST	endoscope	[8]
LARS	IBM	endoscope	[9]
LER	TIMC/IMAG	endoscope	[10]
MKM	Carl Zeiss AG	microscope	[11]
SurgiScope	Elekta	microscope	[12, 13]

The MINOP 2 project included the development of a semi-robotic assistance system to control an extracorporeal digital camera (exoscope) and an endoscope during neurosurgical interventions [14]. To achieve optimal integration into the surgical workflow and space, the implemented user interface enables freehand manipulation of the scopes by the surgeon. The further sections describe the system design, followed by the results of user-centered evaluation with neurosurgeons.

2 Design

The system consists of a control unit, user interface and 2 semi-robotic manipulators, for the exoscope and the endoscope, as shown in Fig. 1-3.

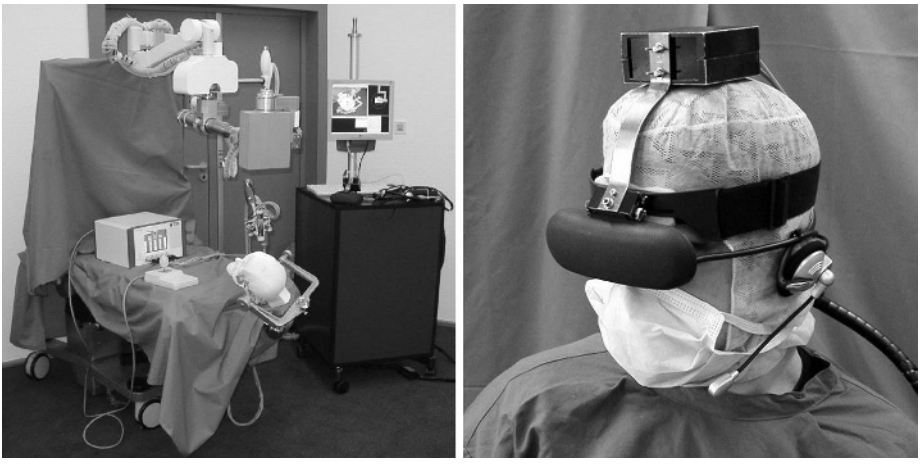


Fig. 1. Labtype of the MINOP 2 system and its user interface

2.1 Control Unit

The control unit of the first labtype system, consisting of a PC, the controller hardware and a flatscreen monitor, is integrated in a trolley, as shown in Fig. 1. It interfaces all other system elements and can be used to transport them.

2.2 User Interface

A head-mounted unit is used by the surgeon during the operation, as shown in Fig. 1. It consists of a binocular display receiving the video image from the exoscope and/or endoscope, an orientation sensor for head tracking, as well as a microphone and earphones for voice control with acoustic feedback.

By using a miniature confirmation button mounted on a finger of the non-dominant hand of the surgeon, an unintentional motion of the manipulators is avoided. This button can alternatively be fixed on or integrated into an instrument such as the suction device. Optionally, the head tracking sensor can be used for manual control as a 3D joystick with integrated confirmation button. In contrast, foot switches were undesirable by the clinical partners, as these are already used for other devices.

Voice control is used to select an active scope and a mode for adjustment, as listed in Tables 2 and 3. Recognized commands are repeated for acknowledgement and performed while the confirmation button is pressed. Depending on the mode, adjustments are proportional to the following head rotation along the transversal, vertical and/or sagittal axis. The active scope is controlled with respect to its “eye” coordinate system, thus ensuring compatibility of view and motion. Releasing the button stops pending adjustments and the user may assume a more comfortable head orientation.

2.3 Exoscope Manipulator

The exoscope is a stereoscopic camera with motorized zoom and focus. It provides two composite video outputs. Illumination of the surgical field is provided by a cold light source through an optical fiber.

The exoscope is mounted on a semi-robotic manipulator, as shown in Fig. 2, utilizing a combination of manual coarse positioning and tele-manipulated robotic fine positioning. The manipulator is fixed on the operation table by a new type of multifunctional arch, as shown in Fig. 1. In case of emergency, the arch can be unlocked and swept aside within a few seconds to provide immediate access to the patient. After fixation, the system can be covered with sterile drape.

Coarse positioning comprises 3 translatory and 2 rotatory (pitch and yaw) degrees of freedom (DOF). The pitch axis is motorized, the other axes are passive and gravity balanced. The pitch and roll axes also possess electromagnetic brakes, which can be released before positioning by pressing a button on a grip. All axes for coarse positioning can be controlled simultaneously with one hand to bring the exoscope directly over the surgical field, typically at the beginning and the end of an intervention, thus allowing a robotic manipulator with compact design and kinematically restricted workspace for increased safety.

Especially during fine-visual preparation work, freehand manipulation of the exoscope is crucial, enabling the surgeon to adjust the system’s position in a more

restricted area, as well as zoom and focus. The robotic fine positioning of the exoscope features 2 translatory and 3 rotatory DOF. After appropriate coarse positioning of the exoscope manipulator, the exoscope's center can be moved on a plane parallel to the surgical field, at a distance currently defined by the neurosurgical partners as 300mm, depending on the length of the endoscopic instruments used. The intersection of the optical axis with that plane is defined as the viewed target.

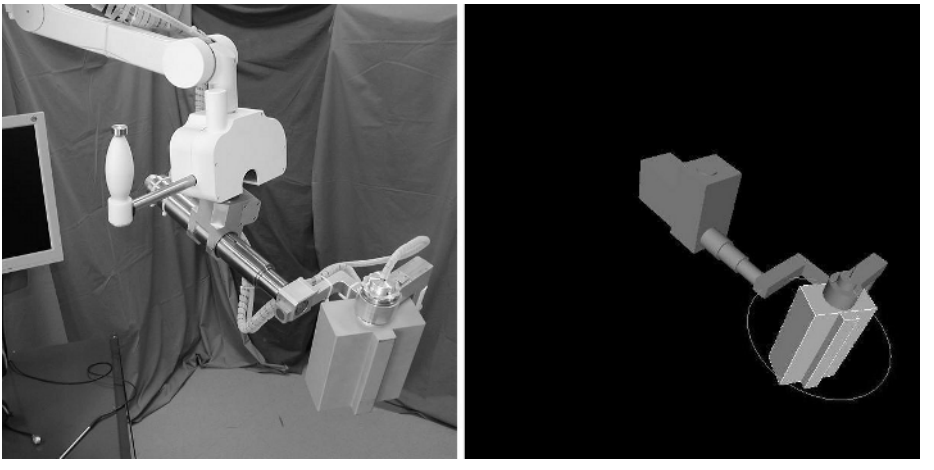


Fig. 2. Robotic exoscope manipulator and its kinematic model

For the safety of the patient and the unobstructed use of endoscopic instruments, fine positioning comprises no DOF allowing translation perpendicular to the surgical field. This confinement is optically compensated by zooming. Also for safety reasons, non-backdrivable motors were used, ensuring instant standstill on power off, while the passive coarse positioning axes allow the manipulator to be swept aside anytime.

Table 2. Exoscope control modes with corresponding adjustments and head rotation

mode	adjustment	transversal	vertical	sagittal
turn	rotation about the exoscope center	pitch	yaw	roll
pivot	rotation about the viewed target	pitch	yaw	roll
lean	rotation about the optical axis			roll
shift	translation par. to the surgical field	up/down	left/right	
zoom	field of view	in/out		
focus	focal distance	in/out		
center	return to center position			
home	find home position and center			

2.4 Endoscope Manipulator

Similar to the exoscope, the endoscope is also mounted on a semi-robotic manipulator. A passive arm with pneumatic brakes is used for fixation on the operation table and coarse positioning.

As shown in Fig. 3, a robot mounted on the arm is used for fine positioning of the endoscope providing 5 DOF. The application specific kinematic design is optimized for minimal obstruction of the operator's view through the exoscope. A parallel kinematic design provides the proximal 4 DOF with high stiffness and low inertia, followed in series by an additional linear spindle drive parallel to the endoscope's optical axis for insertion and retraction. The 6th DOF, i.e. rotation about that axis, will be added in future.

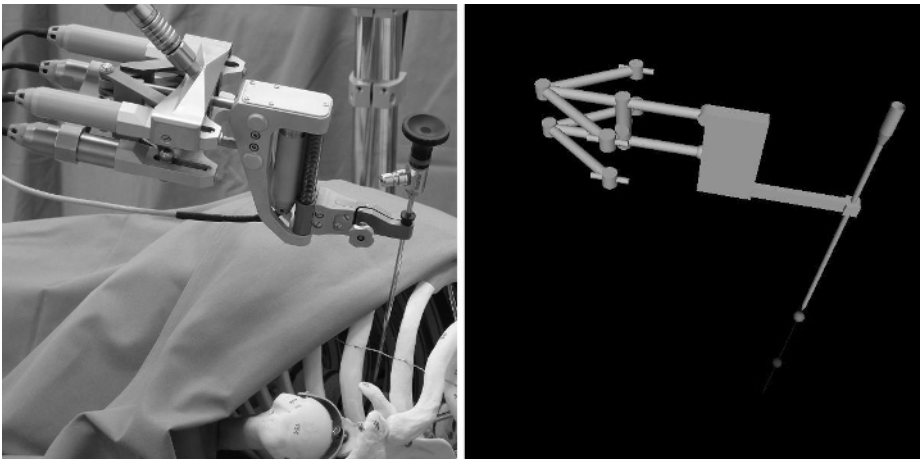


Fig. 3. Robotic endoscope manipulator and its kinematic model

The trocar puncture site through the abdominal wall is defined by software as the remote center of motion. Non-backdrivable motors ensure standstill on an emergency power off. In order to avoid visual and mechanical obstruction by sterile drape, both the holding arm and the robot are autoclavable.

Table 3. Endoscope control modes with corresponding adjustments and head rotation

mode	adjustment	transversal	vertical	sagittal
pivot	rotation about the invariant point	pitch	yaw	roll
lean	rotation about the optical axis			roll
move	translation along the optical axis	forth/back		
zoom	field of view	in/out		
focus	focal distance (if available)	in/out		
center	return to center (insertion) position			
home	find home position and center			

3 Evaluation

The evaluation of the clinical usability of different modes of operation included the rating of effectiveness, efficiency, learnability, reliability and user satisfaction [15].

3.1 User Test in Virtual Reality

At first, a simulation was implemented in stereoscopic virtual reality, as displayed in Fig. 4, to generally evaluate different modes of view control by head rotation. It ran on a standard PC interfacing the head-mounted display, the orientation sensor and the confirmation button. The first test persons were 3 neurosurgeons, 1 surgical assistant and 1 engineer.

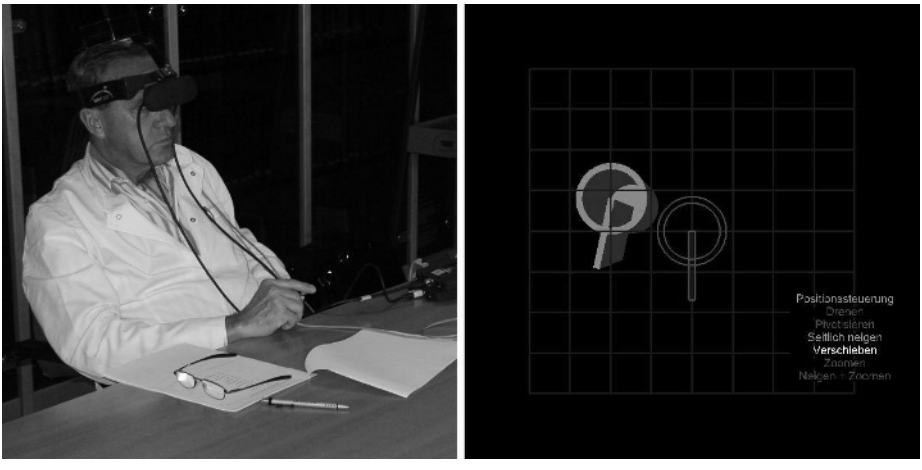


Fig. 4. User test in Virtual Reality

A sequence of pseudo-randomly placed 3D objects was to be targeted by a reticle central to the image, mounted on the virtual scope. Control modes using head tracking (excluding focus) were tested under both position and rate control, i.e. the respective values or their velocity were adjusted proportional to head rotation. The program also measured and documented the time needed to reach the targets. After the test, each user rated the intuitiveness, comfort and controllability of the system using a questionnaire based on rating scales.

The comparison of position and rate control is resumed in Table 4. For each control mode, the mean relative difference was calculated from the measured time (efficiency) and the questionnaire ratings respectively.

The results show the clear advantage of position over rate control. The former was more intuitive, comfortable and controllable for all actions, a tendency increasing with the DOF to be controlled by the operator. A further advantage was the higher efficiency for actions with 3 DOF.

Table 4. Mean relative difference comparing position control to rate control [%]

mode	efficiency	intuitiveness	comfort	controllability
turn	36	16	16	24
pivot	20	13	18	22
lean	-29	7	2	16
shift	-10	11	11	16
zoom	-6	7	7	4

For actions with 1-2 DOF, however, especially lean, rate control proved faster. Users had sometimes to e.g. lean the exoscope more than anatomically possible in one step, thus releasing the button and straightening their neck before continuing in more steps.

In order to keep position control while overcoming its few initial disadvantages, a fast lean function was added. It adjusts the angle multiplied by a customizable factor, hence enabling even large changes in one step. Furthermore, the transmission ratios for shift and zoom were also increased.

3.2 Demonstration on a Phantom

After the implementation of the semi-robotic manipulators and all related soft- and hardware components including the user interface, the system was tested on a phantom together with the clinical partners in an operating room at the Clinic for Neurosurgery of the Mainz University. The initial test results of the system have been encouraging. The users stated that the workspace of the exoscope manipulator can even be reduced, which would automatically lead to a further reduction of its size and weight.

4 Conclusions

We developed an assistance system that enables freehand manipulation of an exoscope and an endoscope by the surgeon, who receives the stereoscopic images and controls the system through a head-mounted unit. The results of the first user tests and demonstration encourage the combination of a head-mounted binocular display, head tracking for analog position control, voice control for mode selection and a miniature confirmation button.

Further optimization of the system and user tests for performance evaluation are objectives of our ongoing work.

Acknowledgements

This research was supported by the German Ministry for Research and Technology (BMBF) within the MINOP 2 project (16SV1442/0).

References

1. Cleary, K., Nguyen, C.: State of the Art in Surgical Robotics: Clinical Applications and Technology Challenges. Computer Aided Surgery, November 2001
2. Taylor, R. H., Stoianovici, D.: Medical Robotics in Computer-Integrated Surgery. IEEE Transactions on Robotics and Automation, Vol. 19, No. 5, October 2003, pp. 765-781
3. Detter, C., Reichenspurner, H., Boehm D.H., Reichart, B.: Robotic manipulators in cardiac surgery: the computer-assisted surgical system ZEUS. Min Invas Ther & Allied Technol 2001: 10(6), pp. 275–281
4. Berkelman, P., Cinquin, P., Troccaz, J., Ayoubi, J., Letoublon, C., Bouchard, F.: A compact, compliant laparoscopic endoscope manipulator. Proc. IEEE Int. Conf. Robotics and Automation, May 2002, pp. 1870–1875
5. Guthart G. S., Kenneth Salisbury, J. J.: The intuitive telesurgery system: overview and application. IEEE International Conference on Robotics and Automation, 2000, pp. 618-621.
6. Aiono, S., Gilbert, J. M., Soin, B., Finlay, P. A., Gordan, A.: Controlled trial of the introduction of a robotic camera assistant (EndoAssist) for laparoscopic cholecystectomy. Surg Endosc (2002) 16:, pp. 1267-1270
7. Buess, G. F., Arezzo, A., Schurr, M. O., Ulmer, F., Fisher, H., Gumb, L. et. al.: A new remote-controlled endoscope positioning system for endoscopic solo surgery. Surg Endosc (2000) 14: pp. 395–399
8. Kim J., Lee Y.-J., Ko, S.-Y., Kwon, D.-S.: Compact Camera Assistant Robot for Minimally Invasive Surgery: KaLAR. Proceedings of 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, September 2004, pp. 2587–2592
9. Taylor, R. H., Fundal, J., Eldridge, B., Gomory, S., Gruben, K., LaRose, D. et. al.: A Telerobotic Assistant for Laparoscopic Surgery. IEEE Engineering in Medicine and Biology, May/June 1995, pp. 279-288
10. Berkelman, P., Boidard, E., Cinquin, P., Troccaz, J.: Control and User Interface Design for Compact Manipulators in Minimally-Invasive Surgery. Proceedings of the 2005 IEEE Conference on Control Applications, August 2005, pp. 25–30
11. Kaminsky, J., Brinker, T., Samii, A., Arango, G., Vorkapic, P., Samii, M.: Technical considerations regarding accuracy of the MKM navigation system. An experimental study on impact factors. Neurol Res 21, 1999, pp. 420-424
12. Lüth, T. C., Hein, A., Albrecht, J., Demirtas, M., Zachow, S., Heissler, E. et. al.: A surgical robotic system for maxillofacial surgery. Proceedings of the 24th Annual Conference of the IEEE Industrial Electronics Society (IECON), 1998, pp. 2470-2475.
13. Sugar, T. G., Fussell, P.: Mouth Operated Neurosurgical Robot. Proceedings of the ASME Design Engineering Technical Conference, September 2001, DETC01/DAC-21130
14. Serefoglou, S., Lauer, W., Perneczky, A., Lutze, T., Radermacher, K.: Multimodal User Interface for a Semi-Robotic Visual Assistance System for Image Guided Neurosurgery. H.U. Lemke et al. (eds.), International Congress Series 1281 (2005), Proc. CARS 2005, pp. 624-629
15. Zimolong, A., Radermacher, K., Stockheim, M., Zimolong, B., Rau, G.: Reliability Analysis and Design in Computer-Assisted Surgery. Universal Access in HCI, Stephanides C, et al (Eds.), Lawrence Erlbaum Ass. Publ., 2003, pp. 524-528 (ISBN: 0-8058-4934-3)

The Feasibility of MR-Image Guided Prostate Biopsy Using Piezoceramic Motors Inside or Near to the Magnet Isocentre

Haytham Elhawary[1], Aleksander Zivanovic[1], Marc Rea[1], Brian Davies[1], Collin Besant[1], Donald McRobbie[2], Nandita de Souza[3], Ian Young[4], and Michael Lampérth[1]

[1] Mechanical Engineering Department, Imperial College London, South Kensington Campus, SW7 2AZ, London, UK
{h.elhawary, a.zivanovic, m.lamperth}@imperial.ac.uk
[2] Faculty of Medicine, Clinical Sciences Centre, Charing Cross Hospital, Imperial College London, W6 8RF, London, UK
[3] Cancer Research UK Clinical Magnetic Resonance Research Group, Institute of Cancer Research, Royal Marsden NHS Foundation Trust, Downs Road, Sutton, Surrey, UK
[4] Electrical Engineering Department, Imperial College London

Abstract. The excellent soft tissue contrast of Magnetic Resonance Imaging (MRI) has encouraged the development of MRI compatible systems capable of combining the advantages of robotic manipulators with high quality anatomical images. Continuing this development, a new five DOF prostate biopsy manipulator has been designed for use inside a closed 1.5T MRI scanner. Space constraints in the bore and the current trend to restrict field strength exposure for operators indicate that a master-slave configuration is ideal for controlling the robotic system from outside the bore. This system has been designed to work with piezoceramic motors and optical encoders placed inside or near the field of view of the scanner, using real time image guidance for targeting biopsies to specific lesions in the prostate. MRI tests have been performed to prove the feasibility of this concept and a one DOF proof-of-concept test rig implementing closed loop position control has been tested and is presented here. A first prototype of the slave manipulator has been designed and manufactured incorporating this new technology.

1 Introduction

1.1 Prostate Cancer

Each year over 27,000 men are diagnosed with prostate cancer in the UK, and it causes approximately 10,000 deaths [1]. Quick and reliable methods for diagnosis are therefore essential for selecting the right type of therapy, if any, to be used in each case. Currently, if a patient is suspected of having prostate cancer, a needle biopsy of the prostate is carried out using trans-rectal ultrasound (TRUS). This technique provides limited image resolution, giving rise to a possibility of diagnostic inaccuracies.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 519 – 526, 2006.

Literature shows that for the sextant biopsy protocol, which is the standard in many centres, TRUS guidance leaves 15-31% of prostate cancers undetected [2]. The good soft tissue contrast of MRI makes it a more favourable means for the detection of lesions in the prostate. A robotic system which can perform accurate biopsies of the prostate under MRI guidance would eliminate many of the existing drawbacks of conventional prostate biopsy procedures.

1.2 MRI Compatible Manipulators

Over the past decade a number of MRI compatible manipulators have been developed [3-7]. The challenges in developing such devices are numerous. Ferromagnetic materials and electromagnetic actuators are prohibited, and noise from electronic circuits can considerably degrade the Signal to Noise Ratio (SNR) of the MR image. In addition, closed MRI scanners impose severe spatial constraints, providing poor access to the patient within the cylindrical bore of the scanner. There is also a growing trend to restrict the field strength to which medical practitioners are exposed [8]. This could give an important impulse to the development of tele-operated MRI compatible manipulators capable of being introduced into the scanner bore.

Two devices which perform MRI compatible prostate biopsy in a closed 1.5T scanner have previously been developed [5, 6], where positioning of the trans-rectal probe is performed by a passive manipulator in both cases. Long flexible rods which control the motion of the manipulator extend outside of the bore, and are actuated manually by the practitioner.

In this paper, a feasibility study is performed to demonstrate the viability of using piezoceramic motors placed inside or very near to the field of view of the MRI scanner to actuate a prostate biopsy manipulator. The motors will be used to position and fire a biopsy needle under real-time image guidance. The feasibility of this concept is demonstrated with a one DOF proof of concept prototype.

2 System Design

2.1 System Requirements and Specifications

The objectives of the system are to use MR image guidance to target abnormalities in the prostate and to perform a biopsy accurately and quickly. The system is designed for use inside a closed 1.5T MRI scanner, and will have a master-slave configuration. An endorectal probe with fiducial markers and an RF coil is inserted into the rectum of the patient, who is in the left lateral decubitus position. The probe is then attached to a slave robotic system that positions it, and fires a biopsy needle into a defined position in the prostate. The operator controls a master manipulator located inside the scanner room, and the slave system located inside the scanner bore, follows the movements of the master. Images from the MRI scanner are obtained in real-time and displayed to the practitioner so that the transrectal probe's position and any displacement of internal tissue can be seen at all times. A block diagram of the system is displayed in Fig. 1.

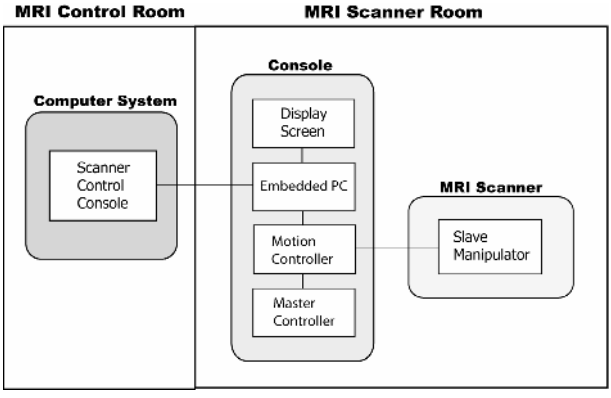


Fig. 1. Block diagram of MRI compatible master-slave prostate biopsy manipulator

To obtain the specifications of the manipulator, a number of conventional TRUS guided prostate biopsy procedures were observed. Movement of the endorectal probe in four degrees of freedom is necessary to reproduce the urologist's movements: three rotations (yaw, pitch and roll) centred on the anus and one translation (in/out). The ranges of motion of the slave manipulator are shown in Fig. 2 *(a)*, together with a kinematic diagram Fig. 2 *(b)*. In addition, a fifth DOF is required to position the biopsy needle (for 2-3cm) at the right depth before firing it into the prostate.

(a) *(b)*

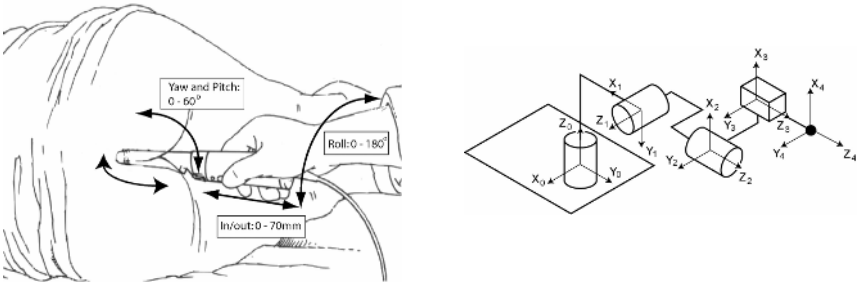


Fig. 2. (a) The four degrees of freedom required to replicate the prostate biopsy procedure, (b) a kinematic diagram of the system

2.2 MRI Compatible Actuation Technologies

In [9] four main categories of motion technologies were identified as suitable for use in MRI compatible manipulators: piezoceramic actuators, pneumatic and hydraulic systems and manual actuation. Piezoceramic motors were selected for use in this prostate biopsy manipulator, as they present precise positioning (in the order of nanometres) and virtually no backlash. They also have a hard brake which makes them rigid even during a power failure, which is an important safety measure for

medical systems. More importantly, they are available in non-magnetic versions and are very compact.

Most of the manipulators that have used piezoceramic motors for actuation have generally followed one of two strategies: *(i)* placing the motors inside the field of view of the scanner – due to noise from the motors interfering with the MR images, this approach requires turning the motors off during image acquisition [7], impeding real-time image guidance – and *(ii)* placing the motors at a distance from the imaging volume in the scanner, to avoid degradation of the Signal to Noise Ratio (SNR) and the creation of image artifacts [3, 4]. A separation of around 1m [4] usually proves enough distance to avoid deterioration of the SNR, but it also means that transmission mechanisms are needed. In [3] long titanium arms were used, while in [4] telescopic acrylic shafts transmitted the motion into the field of view of the scanner. One of the objectives for this system is to incorporate the motors inside or very near the field of view of the scanner, avoiding the need for a transmission mechanism. To achieve this, an alternative model of piezoceramic motors (Piezomotor PiezoLegs) was sourced, with a maximum force of approximately 7N, maximum speed of 12.5mm/sec and good MR compatibility. Small reflective surface mount optical encoders (Agilent AEDR-8300) were sourced and tested, to record the position of the slave.

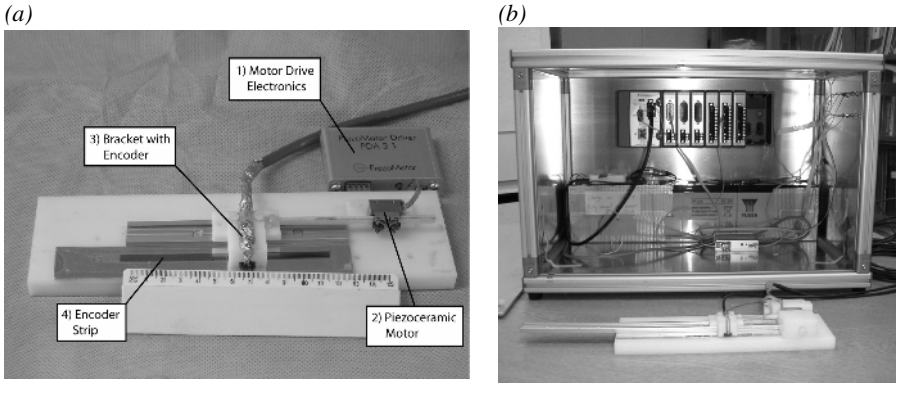


Fig. 3. *(a)* The one DOF prototype with a piezoceramic motor and optical encoder. *(b)* The shielded aluminium enclosure showing the motion controller, batteries and motor electronics.

2.3 Feasibility Studies with a One DOF Test Rig

To demonstrate the feasibility of placing the motors in the field of view of the scanner a one DOF test rig was designed and manufactured, as shown in Fig. 3*(a)*. It included a piezoceramic motor (2) and an encoder (3), to implement closed-loop position control. The rig was constructed with Delrin™ and contained an aluminium rail. A motion controller (National Instruments Compact Rio 9004) with LabView was used to position the motor. For testing in the scanner, a shielded aluminium enclosure was constructed which housed the Compact Rio, a 24V lead acid battery and the motor drive electronics. The enclosure was then connected to the scanner room's ground

connection and placed 1-2m away from the scanner bore, acting as a Faraday cage. All wires that left the enclosure were shielded twisted pair cables, and were filtered with a low pass filter (TUSONIX 4209-053).

3 MRI Tests and Results

MRI compatibility tests were performed with a piezoceramic motor, an encoder and the one DOF test rig. Each component was placed on a fluid filled phantom and scanned in a 1.5T Siemens Magnetom Vision. The image sequences used to test the components were gradient echoes, which are more representative of the real-time nature of the image guided biopsy procedure for which the slave manipulator is designed. There are two principal types of image degradation: *(i)* an artifact that appears on the scan and *(ii)* a degradation of the SNR of the image. If an artifact was detected, then its size was quantified. This indicated how far a certain component had to be placed away from the region of interest to avoid unacceptable distortion of a desired area within the image. With respect to the SNR degradation, the SNR of an image of the component on a phantom was compared to the SNR of an image of the phantom without the component, to quantify the interference produced by the device. The definition of SNR used is the ratio between the mean signal in the 40x40 pixel region in the centre of the phantom on the image, and the standard deviation of the signal in the 40x40 pixel region at the bottom right hand corner of the image [3].

3.1 Encoders and Motors

The encoders were tested with a FLASH 2D gradient echo sequence (FOV=300mm, TR/TE=200/10ms, 256x256, BW=130Hz/pixel, FA: 30º, 9 slices, slice thickness and spacing: 5mm). A single encoder (5.12 x 3.96 x 1.63mm) produced a localised image void enclosed in a rectangular region of 43 x 16mm in the transverse direction. Outside this region, the encoder has no effect on the image.

To test the variation of the SNR produced by the encoder, it was tested under various conditions (unwired, wired and powered) to see its effect on the image. The results for the slice located nearest to the centre of the field of view are presented in Table 1. The SNR of the images with the encoder are compared to the SNR of the phantom alone, and the reduction in SNR is calculated as a percentage. The results indicate that the reduction in SNR due to the encoder is virtually negligible.

Table 1. SNR variation of encoder under different conditions

Elements tested	SNR	SNR reduction (%)
Phantom	138.9	-
Phantom with encoder unwired	138.3	0.4
Phantom with encoder wired but not powered	136.5	1.7
Phantom with encoder powered	135.4	2.6

A non-magnetic linear actuator was tested on a phantom with a True Fisp gradient echo sequence (FOV=230mm, TR/TE: 6.46/3.05ms, 256x256, FA: 80º, 6 slices, slice

thickness and spacing: 5mm). The motor (22 x 10.8 x 18mm) produced its biggest artifact in the transverse direction spanning a rectangular region of 51 x 21mm represented in Fig. 4. Outside this region, the motor does not distort the image.

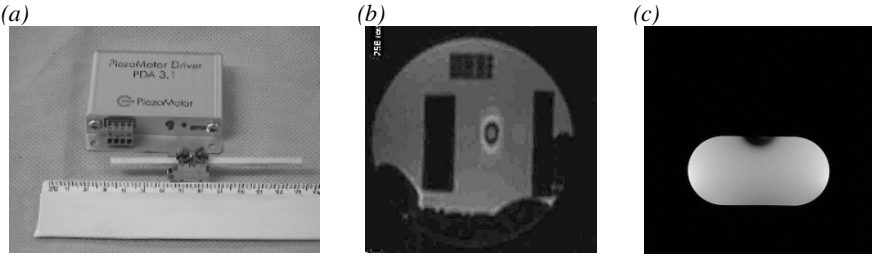


Fig. 4. *(a)* Piezomotor with drive electronics, *(b)* MR images of the phantom with the motor in coronal orientation and *(c)* in transverse orientation

The motor was tested under various conditions to see its effect on the SNR, with the results for a slice located near to the field of view presented in Table 2. The SNR reduction is under 14%, which still makes the image perfectly usable for needle guidance.

Table 2. SNR variation of motor under different conditions

Elements tested	SNR	SNR reduction (%)
Phantom	52.3	-
Phantom with motor unwired	48.7	6.9
Phantom with motor wired but not powered	48.3	7.6
Phantom with motor powered but not actuated	46.2	11.7
Phantom with motor actuated at half max speed	45.0	13.8
Phantom with motor actuated at max speed	45.1	13.7

3.2 One DOF Test Rig

The one DOF test rig was placed on top of a phantom and introduced into the centre of the field of view, with closed loop position control implemented. Scans were taken using a True Fisp sequence (FOV=230mm, TR/TE=6.46/3.05ms, 256x256, FA: 80º, 10 slices, slice thickness and spacing: 5mm). The images only displayed small artifacts on certain slices (the test rig places the motor and encoder 20mm away from the phantom). The SNR was tested under various conditions with the results for the slice nearest to the field of view presented in Table 3. The SNR degradation is greater than that of the motor and encoder considered separately, possibly due to the fact that the encoder produces pulses as it passes the lines on the encoder strip, which have some harmonics that can interfere with the sensitive RF coils of the scanner. This effect can be reduced by increasing the attenuation of the filters in the shielded enclosure at the Larmor frequency of the scanner. This feature is currently being implemented and this should result in a reduction of noise levels. In addition, the motors and encoders on

the slave manipulator will be located several cm away from the main region of interest, which should also greatly reduce the noise interference produced.

Table 3. SNR variation of 1 DOF testrig under different conditions

Elements tested	SNR	SNR reduction (%)
Phantom	63.2	-
Phantom with test rig wired but not powered	61.8	2.2
Phantom with test rig powered	46.3	26.7
Phantom with test rig with position control	45.7	27.6

4 Hardware Developments

Once the feasibility of implementing closed loop position control with real time image guidance had been demonstrated, a first prototype of the slave manipulator incorporating this technology was built. This prototype incorporates only 2 DOF in an XY stage and is shown in Fig. 5.

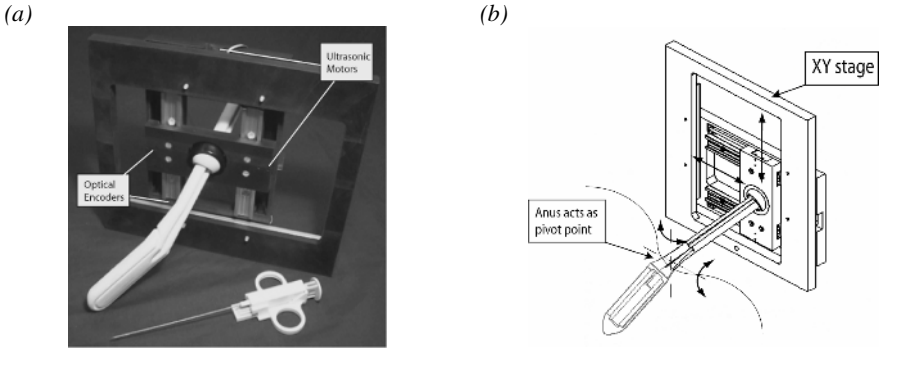


Fig. 5. Slave manipulator with endorectal probe and MRI compatible biopsy needle. The piezoceramic motors and encoders are incorporated into the structure.

An endorectal probe is placed into the anus, with the sphincter acting as a pivot point. The slave, which is made of glass filled Nylon, moves the other extreme of the probe (which is hinged via a plastic spherical bearing) varying the pitch and yaw angles. Each axis contains two piezoceramic motors which are embedded in the structure, two aluminium rails to provide smooth linear motion and an optical encoder. The slave is attached to the scanner table via a positioning arm.

Tests in the MRI scanner have proven that this manipulator is compatible, although rigorous tests to quantify the impact on SNR are still pending. Further design is underway to incorporate the other three DOF: probe translation, roll and needle depth insertion. Work is ongoing to interface the master console, the graphical user interface of the system and the MRI scanner to provide a fully integrated system. Target accuracy must also be assessed on phantoms before moving onto clinical trials.

5 Conclusions

The master-slave prostate biopsy system described in this paper intends to introduce piezoceramic motors into the field of view of the scanner, eliminating the need for a transmission mechanism. Feasibility tests have been performed with a one DOF test rig to see if this objective is viable, yielding positive results. The test rig, which included an encoder and a piezoceramic motor, implemented closed loop position control with a reduction of under 28% of the SNR and 36μm accuracy. This renders images with sufficient quality for MR image guided biopsy. The power source and electronics for the system were housed in an aluminium shielded enclosure placed just outside the entrance of the scanner bore, and all cables exiting the enclosure were filtered. The noise can be further reduced by improving the attenuation of these filters and tests are currently underway to demonstrate this. A first prototype of the slave manipulator has been designed and manufactured which includes 2 of the 5 DOF required.

References

1. Cancer Research UK, Men's Cancers Factsheet - June 2004, World Wide Web: http://info.cancerresearchuk.org/images/publicationspdfs/factsheet_men_June2004.pdf
2. D'Amico AV, Tempany CM, Cormack R, Hata N, Jinzaki M, Tuncali K, et al. Transperineal magnetic resonance image guided prostate biopsy. J Urol 2000;164(2): pp. 385-7.
3. Chinzei K, Hata N, Jolesz FA, Kikinis R. Surgical Assist Robot for the Active Navigation in the Intraoperative MRI: Hardware Design Issues. In: Proc. 2000 IEEE/RSJ International Conf Intelligent Robots and Systems; 2000; Maui, HI, USA: 1. pp. 727-732.
4. Larson BT, Erdman AG, Tsekos NV, Yacoub E, Tsekos PV, Koutlas IG. Design of an MRI-Compatible Robotic Stereotactic Device for Minimally Invasive Interventions in the Breast. Journal of Biomechanical Engineering - Transactions of the ASME 2004;126(4): pp. 458-465.
5. Beyersdorff D, Winkel A, Hamm B, Lenk S, Loening SA, Taupitz M. MR imaging-guided prostate biopsy with a closed MR unit at 1.5 T: initial results. Radiology 2005;234(2): pp. 576-81. Epub 2004 Dec 22.
6. Fichtinger G, Krieger A, Susil RC, Menard C, Coleman JA, Atalar E, et al. Transrectal Prostate Biopsy Inside Closed MRI Scanner with Remote Actuation, under Real-Time Image Guidance. In: Medical Image Computing and Computer-Assisted Intervention; 2002; Tokyo, Japan: Lecture Notes of Computer Science 2488, Part 1. pp. 91 - 98.
7. Masamune K, Kobayashi E, Masutani Y, Suzuki M, Dohi T, Iseki H, et al. Development of an MRI-compatible needle insertion manipulator for stereotactic neurosurgery. Journal of Image Guided Surgery, Copyright © 1995;1(4): pp. 242 - 248.
8. Directive 2004/40/EC of the European Parliament and of the council of 29 April 2004 on the minimum health and safety requirements regarding the exposure of workers to the risks arising from physical agents (electromagnetic fields),
9. Elhawary H, Zivanovic A, Davies B, Lamperth M. A Review of Magnetic Resonance Imaging Compatible Manipulators in Surgery. Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine 2006;220(3): pp. 413-424.

The Role of Insertion Points in the Detection and Positioning of Instruments in Laparoscopy for Robotic Tasks

Christophe Doignon, Florent Nageotte, and Michel de Mathelin

Control, Vision and Robotic Team,
LSIIT (UMR ULP-CNRS 7005) University of Strasbourg, France
name@lsiit.u-strasbg.fr

Abstract. In robot-assisted laparoscopic surgery, an endoscopic camera is used to control the motion of surgical instruments. With this minimally invasive surgical (MIS) technique, every instrument has to pass through an insertion point in the abdominal wall and is mounted on the end-effector of a surgical robot which can be controlled by visual feedback. To achieve an accurate vision-based positioning of laparoscopic instruments, we introduce the motion constraint in MIS which is based on the location of out-of-field of view insertion points. The knowledge of the (image of the) insertion point location is helpful for real-time image segmentation issues, particularly to initiate the search for region seeds corresponding to the instruments. Moreover, with this "eye-to-hand" robot vision system, visual servoing is a very convenient technique to automatically guide an instrument but it requires the velocity screw to be expressed in the appropriate frame. Then, the location of the insertion point is seen as the main part of the larger problem of determining the overall transformation between the camera and the robot end-effector frame. This is achieved thanks to a novel algorithm for the pose determination of cylindrical-shaped instruments. With the proposed method, the location of insertion points can be recovered, on-line, with no marker, without any knowledge of robot kinematics and without an external measurement device.

1 Introduction and Motivations

In laparoscopic surgery, motions of surgical instruments are constrained to by the insertion point locations in the abdominal wall [1], reducing the mobility since only 4 degrees of freedom are available (see Fig. 1 (left)). Endoscopic vision systems may be used for intra-operative guidance of instruments, however to achieve an accurate positioning inside the abdominal cavity, many hard problems must be overcome. Well-known difficulties are emanating from the time-varying lighting conditions, the presence of specularities, bloodstained parts, smoke and a non-uniform and moving background due to patient breathing and heart beating. The major problem for this "eye-to-hands" robot vision system is the unknown position/orientation of the instrument w.r.t. the camera frame which must be recovered to properly guide instruments thanks to visual servoing techniques. Visual servoing consists of using the data provided by cameras to

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 527–534, 2006.

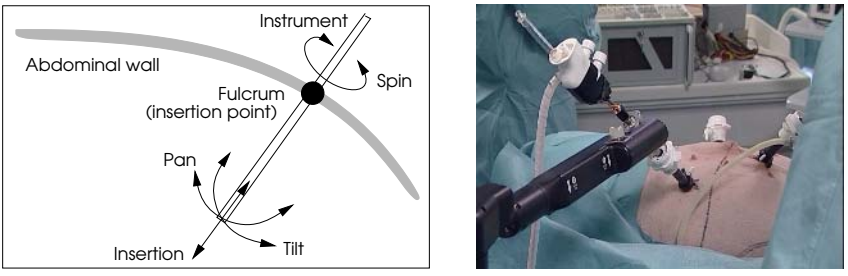


Fig. 1. (left) The free 4 DOFs (insertion/pan/tilt/spin) in laparoscopic surgery. (right) The instrument mounted onto the AESOP surgical robot in the operating room.

control the motion of a robot [2]. To this aim, a set of visual features $\mathbf{s}$ must be designed from the data measurements to control the desired degrees of freedom (DOF). To perform an image-based visual servoing, one needs to compute an interaction matrix $\mathtt{L}_s$, which links the time variations of visual features to the relative camera-object kinematics screw $\boldsymbol{\tau}$ which is defined by $\dot{\mathbf{s}} = \mathtt{L}_\mathbf{s}\ \boldsymbol{\tau}$ [3]. To control the robot using the joint velocities, we must have $\dot{\mathbf{s}} = \mathtt{J}_s\ \dot{\mathbf{q}} + \frac{\partial \mathbf{s}}{\partial t}$, and $\mathtt{J}_\mathbf{s} = -\mathtt{L}_\mathbf{s}\ {}^c\mathtt{T}_i\ {}^i\mathtt{T}_e\ {}^e\mathtt{J}(\mathbf{q})$, where ${}^e\mathtt{J}(\mathbf{q})$ is the robot Jacobian expressed in the end-effector frame (R_e). ${}^c\mathtt{T}_i$ is the time-varying transformation between the camera frame (R_c) and the instrument frame (R_I). With the 4 DOFs available in MIS, if $\mathbf{v}$ is the vector corresponding to the penetration direction, it is expressed with

$$ {}^c\mathtt{T}_i = \begin{bmatrix} \mathbf{v} & [\mathbf{t}]_\times\ \mathtt{R} \\ \mathbf{0}_3 & \mathtt{R} \end{bmatrix} , \tag{1} $$

where $\mathtt{R} = (\mathbf{r}_1, \mathbf{r}_2, \mathbf{v})$ is the rotation matrix and $\mathbf{t}$ is the translation vector between the camera frame (R_c) and the instrument frame (R_I). The estimation of vectors $\mathbf{v}$ and $\mathbf{t}$ is the main objective and contribution of the paper. The organization is as follows. In the next section, we review some endoscopic vision systems used in robot-assisted laparoscopy. In section three, we briefly present the image segmentation, we formalize the motion constraint together with the 3D pose of cylindrical-shaped objects. In section four, we present experimental results for the location of the insertion point in surgical conditions.

2 Related Work on Vision-Based Robotic Guidance for Minimally Invasive Abdominal Surgery

Prior works have been conducted to process laparoscopic images for the development of navigation systems in the human body. Wei *et al.* [4] have used a laparoscope mounted on a robot arm and have designed a color marker to realize a tracking task. This spectral mark was then utilized to control the robot motion at a sampling rate of 15 Hz. Wang *et al.* [5] have proposed to enhance laparoscope manoeuvering capabilities. In so doing, they have conceived a general framework to assist the surgeon in manipulating a laparoscope mounted on

a robot end-effector. This framework has been applied to the 2D positioning of the imaged tip of instrument with the AESOP robot in a way to follow the laparoscope. Like for the previous related work, it's a 2D visual tracking system.It is also assumed that the camera is mounted on a robot (eye-in-hand). Other more recent works are rather related to the 3D tracking with a stationary camera. Hayashibe *et al.* [6] have designed an active scanning system with structured lighting for the reconstruction of 3D intraoperative local geometry of pointed organs. A robot vision system that automatically positions a single laparoscopic instrument with a stationary camera is described by Krupa *et al.* [7]. A laser pointing device has been designed to emit markers on the organ. A visual servoing algorithm is carried out to position a marked instrument by combining pixel coordinates of the laser spots and the estimated distance between the pointed organ surface and the tip of the instrument. It is worth noticing that a on-line identification of the Jacobian matrix for pan/tilt control was realized with robot joint motions to directly get expressions of the velocity screw in the instrument frame. At the Center for Computer Integrated Surgical Systems and Technology (CISST), several techniques for assisting surgeons have been developed. Many of them involve (mono- and stereo-) vision-based robot control with articulated instruments [8]. To identify the robot (fixed) frame-to-camera transformation, the Optrotrak system is used. Burschka *et al.* have noticed an offset of approximately 5 mm which is due to the difficulty of segmenting led centers.

Some solutions for the recovery of insertion points have been provided by Krupa *et al.* [7] and also by Ortmaier *et al.* [9] but with respect to the robot frame, which inherently introduces errors of the robot model. Moreover, these methods need specific markers. Robotic tasks may require interactions with tissues, instruments must be autoclavable before a surgical operation and since several one may alternatively be used (depending on the subtask addressed), it is not convenient to always use artificial landmarks placed on endoscopic tools.

3 Localization of Instruments and Motion Constraint

In this section, we formalize the motion constraint. A scene structure from motion approach is developed to get the location of insertion points. To this aim, the automatic segmentation of instruments is briefly presented and a novel algorithm for the 3D pose estimation of cylindrical-shaped instrument is detailled.

3.1 The Motion Constraint in MIS

The multiview approach we present exploits existing motion constraints of instruments observed by a stationary camera. We denote with (R_c) the camera frame with projection centre $\mathbf{C}$, (R_I) the reference frame attached to a laparoscopic instrument with an arbitrary origin $\mathbf{O}_I$. We define a unit vector $\mathbf{v}$ as the direction of instrument axis. The small incision area in the abdominal wall is represented with a geometrical point $\mathbf{I}$ (Fig. 2 (left)). With the notations used in (1), we have $\mathbf{CI} = \mathbf{t} + \mathtt{R}\,\mathbf{O}^I\mathbf{I} = \mathbf{t} + \lambda\,\mathtt{R}\left[0\,0\,1\right]^{\mathsf{T}} = \mathbf{t} + \lambda\,\mathbf{v}$, for any

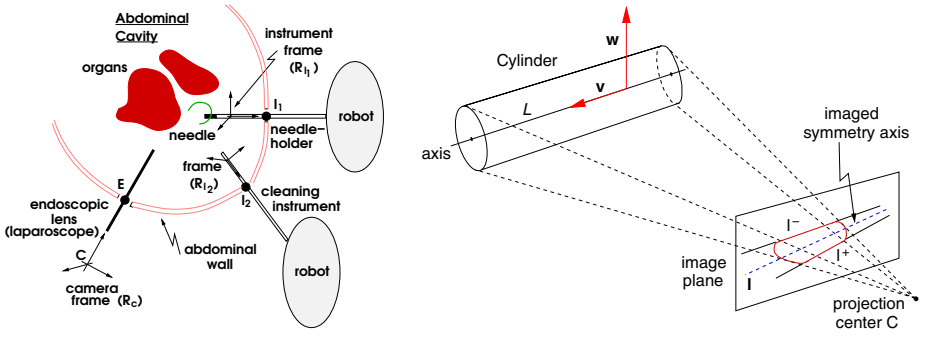


Fig. 2. (left) In laparoscopy, endoscopic camera can observe organs and parts of instruments with motions constrained to insertion points $I_1, I_2, ...$ on the abdominal wall. (right) Plücker coordinates of the cylinder axis and its perspective projection.

scalar $\lambda \in \mathbb{R}$. Since most instruments exhibit a surface of revolution (SOR), the attitude of the axis of revolution may conveniently be represented with the Plücker coordinates as it is for any 3D straight line L. It is a couple of algebraically dependent vectors $(\mathbf{v}, \mathbf{w})$ such that $\mathbf{v}^\mathsf{T}\mathbf{w} = 0$. Thus $\mathbf{w}$ is chosen as the vector normal to the interpretation plane of the viewed line, that is the plane passing through the projection center and the 3D line (see Fig. 2 (right)). They may alternatively be gathered in the following (singular) matrix $\mathtt{L}$ or its dual $\mathtt{L}^\star$ ($\mathtt{L}\,\mathtt{L}^\star = \mathtt{0}$):

$$\mathtt{L} = \begin{bmatrix} [\mathbf{w}]_\times & -\mathbf{v} \\ \mathbf{v}^T & 0 \end{bmatrix} \quad , \quad \mathtt{L}^\star = \begin{bmatrix} [\mathbf{v}]_\times & -\mathbf{w} \\ \mathbf{w}^T & 0 \end{bmatrix} \ . \tag{2}$$

The motion constraint can be expressed as the intersection of multiple convergent 3D straight lines. Since any (homogeneous) point $\mathbf{X}$ is on L if $\mathtt{L}^\star\mathbf{X} = \mathbf{0}$, given n displacements corresponding to the set of dual Plücker matrices $\{\mathtt{L}_1^\star, \mathtt{L}_2^\star, ..., \mathtt{L}_n^\star\}$, the intersection of lines is obtained with a rank-3 $(4n \times 4)$ matrix $\mathtt{G}_n^\mathsf{T}$ such that $\mathtt{G}_n = [\mathtt{L}_1^\star, \mathtt{L}_2^\star, ..., \mathtt{L}_n^\star]$. That is, the null-space of $\mathtt{G}_n^\mathsf{T}$ must be a one-dimensional subspace and the intersection may be computed with n ($n \geq 2$) 3D displacements. By computing the SVD of $\mathtt{G}_n^\mathsf{T}$, one obtains the common intersection with the singular vector associated to the null singular value. Moreover, the perspective projection of the 3D line L_j is the image line $\mathbf{l}_j$ defined by (see [10], p. 186)

$$[\mathbf{l}_j]_\times = \mathtt{K}^c\mathtt{P}^c\ \mathtt{L}_j\ (\mathtt{K}^c\mathtt{P}^c)^\mathsf{T} \ = \ [(\mathtt{K}^c)^{-\mathsf{T}}\ \mathbf{w}_j]_\times \quad \Rightarrow \quad \mathbf{l}_j \equiv (\mathtt{K}^c)^{-\mathsf{T}}\ \mathbf{w}_j \ , \tag{3}$$

where $\mathtt{K}^c$ is the matrix of camera parameters and $\mathtt{P}^c$ is the projection matrix. Since vector $\mathbf{l}_j$ is defined up to a scale, it does not depend on the magnitude of vector $\mathbf{w}_j$, hence the n convergent image lines $\mathbf{l}_1, \mathbf{l}_2, ..., \mathbf{l}_n$ must satisfy

$$\left(\mathbf{l}_1\ \mathbf{l}_2\ \ldots\ \mathbf{l}_n\right)^\mathsf{T}\ i \ = \underbrace{\left(\mathbf{w}_1\ \mathbf{w}_2\ \ldots\ \mathbf{w}_n\right)^\mathsf{T}}_{\mathtt{W}_n}\ (\mathtt{K}^c)^{-1}\ i \ = \ \mathbf{0} \tag{4}$$

where i is the image of the insertion point I. It follows that a set of n 3D straight lines is projected to n convergent image lines if the above $(n \times 3)$ matrix $\mathtt{W}_n$ is

of rank 2. It's only a necessary condition which does not ensure the convergence of the 3D lines, but which makes so important the accurate estimation of the imaged cylinder axis (lines $\mathbf{l}_j$), hence the estimation of its Plücker coordinates.

3.2 Real-Time Segmentation in the Abdominal Cavity

In laparoscopic surgery, many surgical instruments have metallic parts leading to grey regions in the image. The color image segmentation we designed is based on chromatic HS (Hue-Saturation) color attributes. The joint color feature $\mathcal{H} = S\ H$ is a powerful discriminant cue as shown in Fig. 3 (middle). To get out an oversegmentation, a fast non linear filtering has been especially designed. A region-based segmentation approach has been developed, and since any instrument is constrained to pass through the insertion point, the automatic detection of seeds to initiate a region growing process is reduced to a one-dimensional search of low $\mathcal{H}$ values along the image boundaries. The region boundary-pixels are then extracted and classified in two classes and a robust least median of squares method is used for modelling the apparent contour of cylindrical instrument with a pair of lines ($\mathbf{l}^-$ and $\mathbf{l}^+$, Fig. 3 (bottom)), or to reject the region if the distance between boundary-pixels and the corresponding (putative) line is too large (see [11] for details).

3.3 Pose Computation of a Cylinder

We present a novel algorithm for the pose estimation of a cylinder. As a close related work, Wong *et al.* [12] have proposed an algorithm which is able to recover the orientation and the depth while an image rectification is performed to coincide the imaged revolution axis of a SOR with one image axis and when the image of a latitude circle is available. The method we propose here is designed

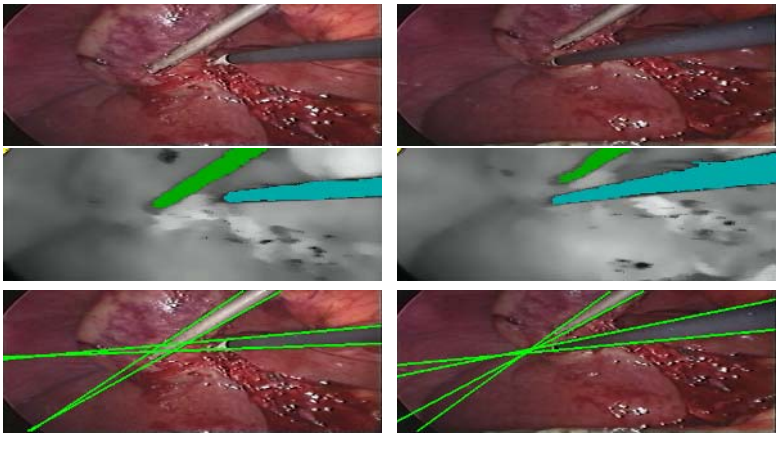


Fig. 3. Segmentation of surgical instruments in the abdominal cavity for two endoscopic images. (top) Original images. (middle) Selected regions in the HS-image. (bottom) Estimation of the pair of straight lines from the apparent contours (in green).

for straight circular cylinders. All components are computed in one stage and it does not need any image transformation and no latitude circle, hence it can deal with partial occlusion of the apparent contour as it is for this application area. Given the radius r_c and the apparent contour, we look for the Plücker coordinates $(\mathbf{v}, \mathbf{w})$ of the cylinder axis. It can be shown with the perspective model that any point m lying on the apparent lines $\mathbf{l}^-$ and $\mathbf{l}^+$ must satisfy

$$\begin{aligned} (\mathbf{l}^-)^\mathsf{T} \; m &\equiv \; \{(\mathtt{K}^c)^{-\mathsf{T}} \; (\mathtt{I} - \alpha[\mathbf{v}]_\times) \; \mathbf{w}\}^\mathsf{T} \; m = \; 0 \\ (\mathbf{l}^+)^\mathsf{T} \; m &\equiv \; \{(\mathtt{K}^c)^{-\mathsf{T}} \; (\mathtt{I} + \alpha[\mathbf{v}]_\times) \; \mathbf{w}\}^\mathsf{T} \; m = \; 0 \quad , \end{aligned} \tag{5}$$

with $\alpha = r_c / \sqrt{\|\mathbf{w}\|^2 - r_c^2}$. To compute the pose parameters, we define vectors $\mathbf{y} = \alpha[\mathbf{v}]_\times \; \mathbf{w}$, $\boldsymbol{\rho}^- = (\mathtt{K}^c)^\mathsf{T} \; \mathbf{l}^-$ and $\boldsymbol{\rho}^+ = (\mathtt{K}^c)^\mathsf{T} \; \mathbf{l}^+$. With these notations, (5) becomes

$$\begin{bmatrix} -\mathtt{I} & \mathtt{I} & -\boldsymbol{\rho}^- & \mathbf{0} \\ \mathtt{I} & \mathtt{I} & \mathbf{0} & -\boldsymbol{\rho}^+ \end{bmatrix} \begin{bmatrix} \mathbf{y}^\mathsf{T} & \mathbf{w}^\mathsf{T} & \mu_1 & \mu_2 \end{bmatrix}^\mathsf{T} = \mathtt{A}_{6\times 8} \;\; \mathbf{x} = \mathbf{0} \tag{6}$$

with the unkwown vector $\mathbf{x} = (\mathbf{y}^\mathsf{T}, \mathbf{w}^\mathsf{T}, \mu_1, \mu_2)^\mathsf{T}$. and μ_1 and μ_2 are two non-null scalars. $\mathbf{y}$ and $\mathbf{w}$ are algebraically dependent since they satisfy $\mathbf{y}^\mathsf{T}\mathbf{w} = 0$ and $\|\mathbf{y}\| = |\alpha| \; \|\mathbf{w}\|$. The latter equation should be developed so as to extract the expression for α

$$r_c^2 \; (\|\mathbf{w}\|^2 \; + \; \|\mathbf{y}\|^2) \; = \; \|\mathbf{w}\|^2 \;\; \|\mathbf{y}\|^2 \tag{7}$$

Since $\mathtt{A}_{6\times 8}$ has a rank equal to 6, the SVD $\mathtt{U}_{6\times 8} \; \mathtt{D} \; (\mathbf{v}_1, \cdots, \mathbf{v}_8)^\mathsf{T}$ has two null singular values and the null-space of $\mathtt{A}_{6\times 8}$ is spanned by the right singular vectors $\mathbf{v}_7$ and $\mathbf{v}_8$ and provides a 2-parameter family of solutions as a linear combination of the two last columns of $\mathtt{V}$ as $\mathbf{x} \; = \; \lambda \; \mathbf{v}_7 \; + \; \tau \; \mathbf{v}_8 \; , \quad \text{for} \;\; \lambda, \tau \in \mathbb{R}$. Then, substituting $\mathbf{y} = (x_1, x_2, x_3)^\mathsf{T}$ and $\mathbf{w} = (x_4, x_5, x_6)^\mathsf{T}$ in $\mathbf{y}^\mathsf{T} \; \mathbf{w} = 0$ gives the following homogeneous quadratic equation in λ and τ $a_1 \; \lambda^2 + a_2 \; \lambda\tau + a_3 \; \tau^2 \; = \; 0$, where a_i are scalar functions of $\mathbf{v}_7$ and $\mathbf{v}_8$. Two real solutions for $s = \tau/\lambda$, s^- and s^+, can be computed. Then, reporting these solutions in (7) gives an homogeneous quadratic equation $c_1(s) \; \tau^2 \; + \; c_2(s) \; \tau^4 \; = \; 0$ and the solutions are $\tau \; = \; 0$ (double) and $\tau = \pm\sqrt{-\frac{c_1(s)}{c_2(s)}}$.

4 Experimental Results

With a training box at the lab, the detection and tracking of a surgical instrument has carried out (see Fig. 4 (top)) with the AESOP surgical robot. During the guidance, we noticed some small temporal variations of the image (i_1) of the insertion point I_1 due to errors in the overall segmentation (Fig. 4 (bottom middle)). In Fig. 4 (bottom right), we have reported the dual parameter space of convergent lines $\mathbf{l}_j$ (distance versus angle of image line direction), since a unique intersection of lines must lead to perfectly collinear points (blue bullets).

In the operating room of IRCAD [1] (see Fig. 1 (right)), this method has been validated with a living pig. The precision of the image of the insertion point

[1] Institut de Recherche contre les Cancers de l'Appareil Digestif, Strasbourg, France.

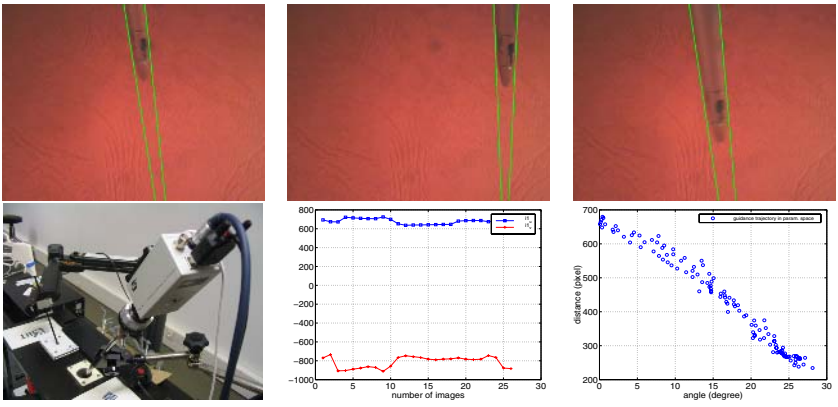


Fig. 4. Experiments in the lab to validate the proposed method. (top) Three images with the detection of a single surgical instrument (in green). (left) A training box is used together with the endoscope fixed onto a monoCCD camera. (middle) Temporal variations of i_1 coordinates in the image while moving the surgical instrument. (right) In the parameter space of convergent lines (θ, ρ) corresponding to the imaged instrument axis, "points" (blue bullets) must be collinear with a motionless insertion point.

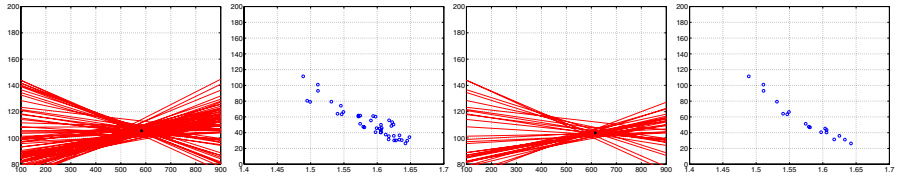


Fig. 5. (left) The convergent imaged symmetric axes and the estimated image of the insertion point i_1 at $(593.4; 105.5)$ (black cross) computed with the least mean squares during the guidance of an instrument in the operating room. (middle left) In the parameter space of onvergent lines (θ, ρ), "points" (blue bullets) must be collinear. (middle right) The estimated image of the insertion point i_1 at $(615.5; 103.9)$ (black cross) and the parameter space (right) with the robust estimation when 50 % of data are rejected.

$i_1 = (593.4; 105.5)$ (see Fig. 5 (left)) is given by the standard deviations $\sigma_u = 25.4$ and $\sigma_v = 4.1$ pixels respectively in the horizontal (u) and vertical directions (v) with the least mean squares method. These results show a significantly better precision in the vertical direction. This can be explained either by the breathing motion or by a no sufficient spread of orientation motions in one direction while the robot is guiding the instrument. By performing a least median of squares method, 50 % of displacements have been discarded (see Fig. 5 (right)) so as to get a robust estimate of i_1. Standard deviations are then $\sigma_u = 11.4$ pixels and $\sigma_v = 2.6$ pixels with the robust method. Finally, the Plücker coordinates of the remaining lines are used to get the 3D location of $I_1 = (-258.9 \pm 29.4; -68.4 \pm 5.2; 316.4 \pm 34.7)$ with the selected displacements.

5 Conclusion

In this paper, we have proposed a method to estimate the location of out-of-field of view insertion points on the abdominal wall suited for vision-based robotic guidance in MIS. This is helpful to drive the image segmentation based on seeded region growing and it is needed for expressing the velocity screw in the appropriate frame so as to track surgical instruments with visual servoing techniques.

For cylindrical-shaped laparoscopic instruments, we have presented a novel algorithm which solves the Plücker coordinates of a cylinder axis and which can deal with partial occlusions of instruments. Despite the disturbance due to the patient breathing, we have shown that this method provides satisfactory results in real surgical conditions with only few endoscopic images.

References

1. Funda, J., Taylor, R., Eldridge, B., Gomory, S., Gruben, K.: Constrained cartesian motion control for teleoperated surgical robots. IEEE Trans. on Robotics and Automation **12** (1996)
2. Hutchinson, S., Hager, G.D., Corke, P.: A tutorial on visual servo control. IEEE Trans. on Robotics and Automation **12** (1996) 651–670
3. Espiau, B., Chaumette, F., Rives, P.: A new approach to visual servoing in robotics. IEEE Transactions on Robotics and Automation **8** (1992) 313–326
4. Wei, G.Q., Arbter, K., Hirzinger, G.: Real-time visual servoing for laparoscopic surgery. IEEE Engineering in Medicine and Biology **16** (1997) 40–45
5. Wang, Y.F., Uecker, D.R., Wang, Y.: A new framework for vision-enabled and robotically assisted minimally invasive surgery. Journal of Computerized Medical Imaging and Graphics **22** (1998) 429–437
6. Hayashibe, M., Nakamura, Y.: Laser-pointing endoscope system for intra-operative geometric registration. In: Proc. of ICRA'01, Seoul, South Korea (2001)
7. Krupa, A., Gangloff, J., Doignon, C., de Mathelin, M., Morel, G., Leroy, J., Soler, L., Marescaux, J.: Autonomous 3-d positioning of surgical instruments in robotized laparoscopic surgery using visual servoing. IEEE Trans. on Robotics and Automation **10** (2003) 842–853
8. Burschka, D., Corso, J.J., Dewan, M., Hager, G.D., Lau, W., Li, M., Lin, H., Marayong, P., Ramey, N.: Navigating inner space: 3-d assistance for minimally invasive surgery. In: Workshop Advances in Robot Vision, IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems, Sendai, Japan (2004) 67–78
9. Ortmaier, T., Hirzinger, G.: Cartesian control issues for minimally invasive robot surgery. In: Proceedings of IROS'00, Takamatsu, Japan (2000)
10. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge Univ. Press (2000)
11. Doignon, C., Graebling, P., de Mathelin, M.: Real-time segmentation of surgical instruments inside the abdominal cavity using a joint hue saturation color feature. Real-Time Imaging **11** (2005) 429–442
12. Wong, K.Y., Mendonça, P.R.S., Cipolla, R.: Reconstruction of surfaces of revolution from single uncalibrated views. Image and Vis. Computing **22** (2004) 829–836

Automatic Localization of Laparoscopic Instruments for the Visual Servoing of an Endoscopic Camera Holder

Sandrine Voros[1], Jean-Alexandre Long[2], and Philippe Cinquin[1]

[1] TIMC-IMAG, UMR CNRS 5525, Université Joseph Fourier, Grenoble
{Sandrine.Voros, Philippe.Cinquin}@imag.fr
[2] Department of Urology, University Hospital, Grenoble

Abstract. The use of a robotized camera holder in laparoscopic surgery allows a surgeon to control the endoscope without the intervention of an assistant. Today, the orders that the surgeon can give to robotized camera holders remain limited. In order to provide higher level interactions between the surgeon and a robotized camera holder, we have developed a new method for the automatic tracking of laparoscopic instruments which works in near real-time. The method is based on the measurement of the 3D positions of the insertion points of the instruments in the abdominal cavity and a simple shape model of the laparoscopic instruments. We present the results of our first experimentation on a cadaver.

1 Introduction

Laparoscopic surgery offers a lot of benefits for the patients compared to open surgery: less blood loss, less transfusions, a smaller consumption of analgesia and a shorter hospitalisation time [1]. However, because of the limited field of view, the specific instruments and the difficult coordination between the assistant manipulating the camera and the surgeon, laparoscopy is a more complex procedure for the surgeon than open surgery. Robotized systems were developed to assist the surgeon in performing laparoscopy, including robotized camera holders: they stabilize the image [2], reduce the staining of the endoscope and enable solo surgery. Several robotized endoscopic holders have been commercialised, including Aesop (Computer Motion) and EndoAssist (Armstrong Healthcare) [3].

Today, the applications allowing the surgeon to control robotized endoscopic holders only integrate simple commands. Providing higher level commands of the robot would be a major clinical added value: it would allow the surgeon to concentrate on his surgical procedure, rather than on the control of the robot. In this objective, the automatic tracking of the instruments is an intuitive approach, since the moving instruments must stay in the field of view of the camera for security reasons.

Several approaches were proposed to detect the instruments in laparoscopic images: [4] designed a robotized instrument holder equipped with a laser beam. The laser beam is detected in the images in order to track the tip of the instrument. This approach raises the questions of cost issues and sterilisation. [5],

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 535–542, 2006.

[6]and [7] taped a color mark on the instrument. The main limit of this approach is the sterilisation of the mark. These four approaches put constraints and add complexity to the procedure. [8] used color classification to identify colors that are likely to correspond to instruments, but the computation time is not indicated. [9] proposed to detect the grey regions in the laparoscopic images, which are likely to correspond to the metallic tip of the instrument. A good computation time is achieved. However, some laparoscopic instruments do not have a metallic tip and the tip might be hidden behind an organ. Finally, [10] proposed an approach based on the detection of the lines in the image, but some parameters have to be tuned by the user. None of these approaches give solutions to identify the instrument that has to be tracked among the other instruments.

Our approach is based on the measurement of the 3D positions of the insertion points of the instruments. With a calibrated robot and a calibrated camera, we can project the insertion point of the instrument we wish to track on the image plane of the camera (even if the insertion point is not visible on the image). This information provides an important constraint on the position of the instrument on the image. It is used to automatically detect the axis, edges and tip of the instrument. The knowledge of the 2D position of the tip allows to perform a visual servoing of the orientation of the camera. For this purpose, we do not require subpixel precision but rather need to roughly determine the position of the tip. As suggested in [7], the orientations of the edges can be used to find the depth of the instrument and include the zoom in the visual servoing loop. [11] showed encouraging results on laparoscopic images and presented our visual servoing on a test bench. In this paper, we focus on the results of our first experimentation on a cadaver.

2 Method

2.1 Calibration of the Robotized Laparoscopic Holder and the Camera

Our robotized camera holder [12] was calibrated, as well as the endoscopic camera [13]: the 3D position of the camera is known in a fixed frame linked to the robot (fig. 1).

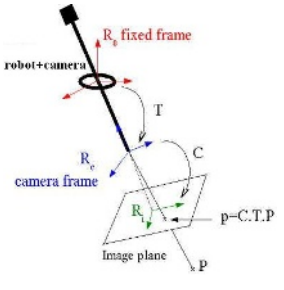


Fig. 1. Calibrated system: if the coordinates of P are known in R_0, its projection on the image plane can be computed thanks to the calibration matrixes T (geometric model of the robot) and C (pinhole model of the camera)

The 3D coordinates of a point can be measured by stereovision: two views of this point for different positions of the robot are sufficient.

2.2 Measurement of the 3D Positions of the Insertion Points of the Instruments

The 3D positions of the insertion points of the instrument are measured at the beginning of the intervention. Since the surgeon gives orders to the robot with a vocal command, we have developed a 'vocal mouse' which allows to move a cursor on the endoscopic image. For two different positions of the camera in which the insertion point is visible, the surgeon selects the position of the insertion point in the image with the vocal mouse.

This initialisation step could be easily integrated to the surgical protocol since the surgeon creates the insertion points at the beginning of the intervention under visual control. We show in section 3 our measurements of the 3D position of an insertion point: the hypothesis that they are relatively fixed during an intervention is validated.

2.3 Model of the Laparoscopic Instruments in the Images

We define $P = proj(T)$ as the projection of an insertion point T in the image plane of the camera. In a laparoscopic image, an instrument can be represented as (fig. 2) :

- three lines: a symmetry axis Δ, and two edges which are symmetrical compared to Δ. The positions of these three lines are constrained by the position of P: $\mathcal{C}$ is a circle centered on P of radius R such that the symmetry axis and the two edges intersect $\mathcal{C}$;
- a point S, which represents the tip of the instrument. S must belong to Δ.

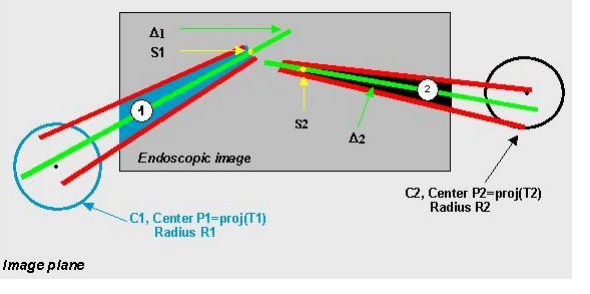
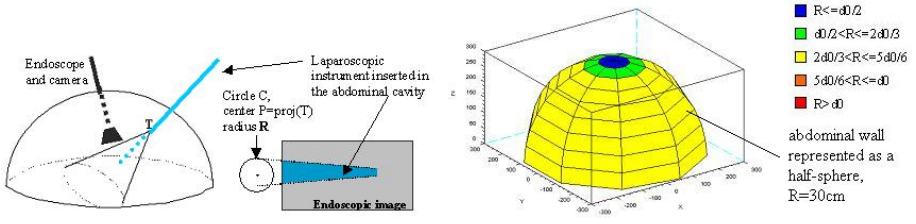


Fig. 2. Representation of the instruments in a laparoscopic image. The position of each instrument in the image is constrained by the position of the projection of its insertion point in the abdominal cavity.

The value of R is determined by a simulation: since the diameter d of a laparoscopic instrument is known (usually 6 mm), we bounded the value of R

from above by the diameter d_0 in pixels the tool would have in the image if it was parallel to the image plane. Our simple model of the abdominal cavity, the instrument and the camera is illustrated by fig. 3, as well as the simulation results.



a) b)

Fig. 3. a) Model of the abdominal cavity, camera and instrument to validate our choice for the value of R, b) result of the simulation: false color representation of R depending on the position of the instrument on the abdominal wall

To detect the instrument inserted through T, we search for a symmetry axis and two edges which intersect $\mathcal{C}$ of radius R, and we search for the tip S of the instrument on the symmetry axis.

2.4 Main Steps of the Detection Algorithm

We suppose that the surgeon has selected an instrument to track (by selecting an insertion point T). The projection P of the insertion point on the image plane is computed.

Step 1: Segmentation
In this step, we compute the edge points in the image by a simple gradient method. The orientation of the edges corresponding to the tool are constrained by the position of P. After this step, we obtain a cloud of edge points, most of which correspond to the instrument.

Step 2: Search for the symmetry axis of the instrument
We use the edge points computed in the previous step to find the symmetry axis of the instrument. To do this, we use a Hough method [14] centered on P. This allows to consider only the lines that have a satisfying orientation, and the computation time is very reasonable compared to a classical Hough method.

Step 3: Search for the edges of the instrument
All the edge points which voted for the symmetry axis were stored in the previous step. The symmetric of the cloud of edge points compared to the symmetry axis is computed to obtain more points, and linear regressions are performed for all the points above the axis and below the axis.

Step 4: Search for tip of the instrument
The tip of the instrument is searched on the symmetry axis. The points of the axis are ordered, according to the position of the insertion point. An Otsu threshold [15] is used to separate the points of the axis in two classes. The pixels of the axis belonging to the same class are grouped into zones. The difficulty of this step is to determine which of the two classes corresponds to the instrument, and to group the zones correctly in order to find the tip. We developped two approach, one for the first image of the tracking, the other for the next images.

For the first image, we search for the longest zone of each class and select the class corresponding to the instrument according to its postion with regard to the tip. The average color of the zone corresponding to the tool and the average color of the zone corresponding to the instrument are stored for the next images. In the next images, we use this color information to determine which zone corresponds to the instrument and to the background.

Once the zone corresponding to the instrument has been found, we look if it can be concatenated with other zones corresponding to the tool according to color and length considerations.

3 Results

The method was tested on anatomical pieces. Our first step was to validate our hypothesis that the insertion points are fixed during the intervention. We measured along time the position of the insertion point of an instrument (section 2.2). To check that the first computation of the insertion point was satisfying, we projected the insertion point on the image plane (fig. 4). The distance between the measured points and the reference was inferior to 5mm except for measure $n°3$, but this was due to leak in of gas in the abdominal cavity. As soon as the leak was stopped the results were satisfying.

$n°$	time (min)	X (mm)	Y (mm)	Z (mm)	distance (mm)
1	0	78.9	148.4	-39.3	reference
2	11	79.2	152.3	-42.3	1.53
3	14	91.6	207	-39.8	57.3
4	25	71.3	143.5	-52.3	4.0
5	43	71	148	-52	1.2
6	205	73.4	147	-53	1.3
7	234	77.14	152.2	-47.6	4.6

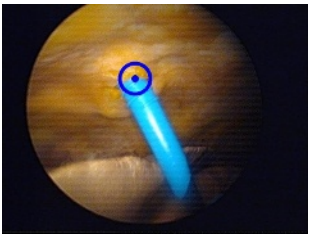


Fig. 4. Variation of the position of the insertion point along time. On the right image, the insertion point was projected on the image plane.

We were able to perform the tracking of an instrument when the specular reflexions were limited. The computation time is around 100 ms for a 200x100 image. It allows a smooth tracking when the instrument is moved at a normal speed. It could still be reduced by optimizing the code allowing to take into

account abrupt and large movements. A resolution of 200x100 pixels is enough for the tracking of a laparoscopic instrument, since we are not looking for subpixel precision to control the camera. We considered that the tool was centered as long as the distance between the detected tip and the previous tip was less than 11 pixels. Fig. 5 shows the distance in pixels between the tip found by the method and the tip manually selected. In 70% of the images, the error is inferior to 5 pixels, and in 87 % the error is less than 11 pixels.

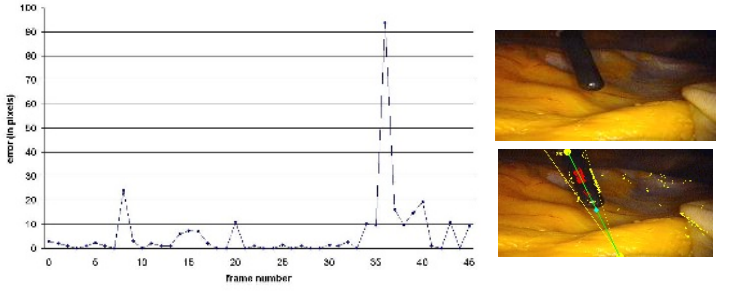


Fig. 5. Error in pixels between the tip found by the method and the tip selected manually and an example of correct detection (yellow dots: candidate points, red points: points corresponding to the axis, blue dot on the axis: tip of the instrument)

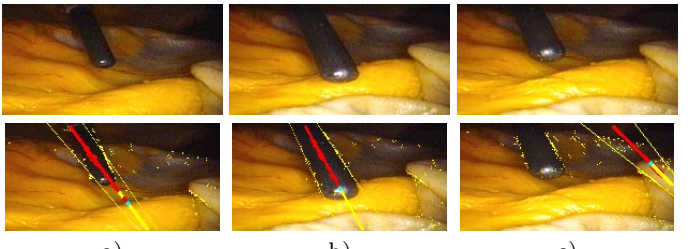


Fig. 6. a) wrong detection due to a lack of contrast b) wrong detection due to specular reflections, c) wrong detection of the axis

We analysed the images for which the detection was wrong (fig. 6): fig. 6 a) corresponds to the first error peak (frame 9). The detection of the axis is correct, but the tip is wrongly positioned: the red segments correspond to the pixels labelled as pixels belonging to the instrument and the yellow segments as pixels labelled as the background. Since the background close to the tip is very dark, it is wrongly labelled by the Otsu threshold as a region belonging to the instrument. This is also the cause of the error for frames 35, 36, 61, 63 and 64.

Fig 6 b) corresponds to the second error peak (frame 21). Again, the axis is correctly detected, but not the tip. This time, the error is due to specular reflexions, which are labelled as background. This is also causing errors for frames 67 and 69. The important error for frame 37 is caused by a false detection of the

axis of the tool (fig. 6 c)): the edges of the instrument are barely visible in the image because of the lack of contrast, and the organ on the far hand side of the image provides a lot of edges, which deteriorates the axis detection.

When the specular reflecions are important, they do not deteriorate the detection of the axis, since they are along the axis of the instrument. However, they cause the failure of tip detection, because the wrong classes are attributed to the instrument and background.

4 Discussion

This experiment allowed us to validate our approach by showing that the insertion point of the instruments are relatively fixed during the intervention. We were able to robustly detect the axis of an instrument, except in cases for which the tool is in a border of the image with a low contrast, or very close to the insertion point. To enhance further this detection, we could give more credit to radial lines than lines with other orientations, since the instruments are usually directed towards the center of the image, as suggested by [10]. The detection of the edges is also quite robust, and we plan to integrate the zoom in the visual servoing of the robot.

The detection of the tip works well on images with little specular reflexions. It must be improved to deal with them. We are currently investigating a solution in which the specular reflexions along the axis of the instrument would be detected and removed. Since we detect correctly the axis, we could perform this detection only on a region of interest around the axis, with a reasonable computation time.

To further enhance the speed of the detection, we also plan to take into account the position of the tip of the instrument in the previous images. However, considering the very strong magnification of an endoscope, the instrument might very well move a lot in the image between two frames, so the region of interest (ROI) used must be carefully chosen. Moreover, if the ROI is too small, only a tiny portion of the instrument will be visible, and our shape model will not be accurate, but we consider making a more complete model of the tool.

Finally, we must avoid as much as possible false detections. Our method deals partly with them: if the amount of edge points which voted for the axis is too low, or if the length of the instrument is too small an error is raised. If the tool was correctly detected in the first image, its color and the background color are stored for the next images. Thus, if the detection algorithm finds a tool which color is very close to the color of the background, an error is raised. We must now find a method to reject a false detection when the image is the first image of the tracking. One solution could be to gather a few images of the abdominal cavity without an instrument at the beginning of the intervention to obtain information about the color of the background.

5 Conclusion

We developed a novel method to detect the instruments in laparoscopic images. The method is automatic and works almost in real-time. Our first cadaver test

showed encouraging results and allowed us to find the difficult cases still to solve. Several high level control modes based on this method could be implemented to supervise surgical tasks such as the insertion/removal of an instrument or suturing tasks.

Acknowledgement. The authors would like to thank Bruno Thibaut from the anatomy laboratory for his help during these experiments.

References

1. B. Makhoul, A. De La Taille, D. Vordos, L. Salomon, P. Sebe *et al.*, Laparoscopic radical nephrectomy for T1 renal cancer: the gold standard? A comparison of laparoscopic vs open nephrectomy. BJU International 2004, Vol. 93, pp. 67-70, 2004.
2. L. R. Kavoussi, R. G. Moore, J. B. Adams et *al.*, Comparison of robotic versus human laparoscopic camera control, J Urol, 154:2134, 1995.
3. P. Ballester, Y. Jain, K. R. Haylett and R. F. McCloy, Comparison of task performance of robotic camera holders EndoAssist and Aesop. International Congress Series, Vol. 1230, pp. 1100-1103, June 2001.
4. A. Krupa, J. Gangloff, C. Doignon, M. de Mathelin, G. Morel *et al.*, Autonomous 3-D positioning of surgical instruments in robotized laparoscopic surgery using visual servoing. IEEE Trans. on Robotics and Automation, vol. 19(5), pp. 842 -853, 2003.
5. G. Wei, K. Arbter, G. Hirzinger, Real-Time Visual Servoing for Laparoscopic Surgery. Controlling Robot Motion with Color Image Segmentation. IEEE Engineering in Medecine and Biology, pp. 40-45, 1997.
6. X. Zhang, S. Payandeh, Application of Visual Tracking for Robotic-Assisted Laparoscopic Surgery, Journal of Robotics Systems (2002) 19(7) : 315-28.
7. O. Tonet, T.U. Ramesh, G. Megali, P. Dario, Image analysis-based approach for localization of endoscopic tools, Proc. of Surgetica'05, pp. 221-228, 2005.
8. Yuang Wang, D. R. Uecker and Yulun Wang, A new framework for vision enabled and robotically assisted minimally invasive surgery. Computerized Medical Imaging and Graphics, Vol. 22, pp. 429-437, 1998.
9. C. Doignon, F. Nageotte, M. De Mathelin, Detection of grey regions in color images : application to the segmentation of a surgical instrument in robotized laparoscopy, Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp 3394-3399, 2004.
10. J. Climent, P. Marés, Automatic instrument localization in laparoscopic surgery, Electronic Letters on Computer Vision and Image Analysis, 4(1), pp. 21-31, 2004
11. S. Voros, E. Orvain, J-A. Long, P. Cinquin, Automatic detection of instruments in laparoscopic images: a first step towards high level command of robotized endoscopic holders, proceedings of the IEEE / RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics, 2006.
12. P. J. Berkelman, Ph. Cinquin, J. Troccaz, J-M. Ayoubi, C. Létoublon, Development of a Compact Cable-Driven Laparoscopic Endoscope Manipulator. MICCAI 2002, Vol. 2488, pp. 17-24, 2002.
13. Z. Zhang, A Flexible New Technique for Camera Calibration. IEEE Trans. on Pattern Analysis and Machine Intelligence, 22(11) pp. 1330-1334, 2000.
14. Richard O. Duda, Peter E. Hart, Use of the Hough Transformation To Detect Lines and Curves in Pictures. Communications of the ACM, Vol. 15(1), pp. 11-15, 1972.
15. N. Otsu, A threshold selection method from gray level histograms. IEEE Trans. Systems, Man and Cybernetics, Vol. 9, pp. 62-66, 1979.

A Novel Robotic Laser Ablation System for Precision Neurosurgery with Intraoperative 5-ALA-Induced PpIX Fluorescence Detection

Masafumi Noguchi[1], Eisuke Aoki[1], Daiki Yoshida, Etsuko Kobayashi[1], Shigeru Omori[2], Yoshihiro Muragaki[3], Hiroshi Iseki[3], Katsushige Nakamura[4], and Ichiro Sakuma[1]

[1] Graduate School of Frontier Sciences, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
{masamasa, aoki, dai, etsuko}@miki.pe.u-tokyo.ac.jp, isakuma@k.u-tokyo.ac.jp
[2] R&D Center, Terumo Corporation, 1500 Inokuchi, Nakai-machi, Ashigarakami-gun, Kanagawa 259-0151, Japan
Shigeru-Oomori@terumo.co.jp
[3] Faculty of Advanced Techno-surgery, Institute of Biomedical Engineering and Science, Graduate School of Medicine, Tokyo Women's Medical University, 8-1 Kawada-cho, Shinjuku-ku, Tokyo 162-8666, Japan
{ymuragaki, hiseki}@abmes.twmu.ac.jp
[4] Mitaka Ko-Ki Corporation, 5-1-4 Osawa, Mitaka-shi, Tokyo 181-0051, Japan

Abstract. We developed a combined system of tumor detection by 5-ALA-induced PpIX fluorescence and precise ablation by micro laser for the first time, with an automatic focusing and robotic scanning mechanism for the brain surface. 5-ALA accumulates on tumors to be metabolized to become PpIX that is a fluorescent. Intra-operative detection of 5-ALA induced PpIX fluorescence provides useful information for tumor detection. The wavelength of the micro laser is 2.8 μm close to the absorption band of water. This laser is effective only on the surface of brain tissue, enabling precise ablation at the boundary between tumor and normal tissue identified by intra-operative 5-ALA induced fluorescence. Combination tests of the fluorescence measurement and the laser ablation were performed, and it was possible to extract the area with fluorescence appropriately from the measurement data, and the micro laser with automatically scanning selectively ablated the extracted area.

1 Introduction

In current neurosurgical practice, surgeons can remove most of a tumor with an accuracy of a few millimeters by using a combination of conventional surgical instruments, such as an electric cautery, and a computer-aided navigation system based on diagnostic images, such as MR and CT images. Residual tumor, especially if the malignant tumor like glioma, may impair the prognosis of the patients and it is necessary to remove as much of the tumor as possible while keeping the normal tissue

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 543 – 550, 2006.

intact. However, it is difficult to know the exact boundary between tumor and normal tissue, and excessive ablation of the normal brain tissue will damage its function.

In craniotomy procedure, deformation of the brain tissue, called "brain shift", occurs due to cerebrospinal fluid leakage and surgical interventions. In some cases, brain shift reaches to several tens of millimeters and continuously increases during the procedure [1]. This requires navigation based on intraoperative MR images. This navigation, however, contains a few millimeters of errors at a maximum caused by a registration of preoperative diagnosis images [1] and intraoperatiove images and an accuracy of the position tracking using such as an optical marker. Furthermore, there is a tradeoff between high frequency of image acquisition for more accurate navigation and not time-consuming imaging.

Similarly, the accuracy of conventional surgical procedures is a few millimeters for removal of residual tumors. Therefore, we desired to achieve a more precise operation with an accuracy of sub-millimeters. To that end, each of a measurement and removal of residual tumors has to realize this accuracy.

To solve these problems, we have proposed a novel approach to therapy using 5-aminolevulinic acid (5-ALA) [2][3] and a micro-laser ablation system [4], with the boundary between the tumor and the normal tissue distinguished by the 5-ALA-induced protoporphyrin IX (PpIX) fluorescence in the tumor and with accurate ablation of the tumor with the micro laser. 5-ALA, which is orally administrated to a patient, accumulates on tumors to be metabolized to become PpIX that is a fluorescent substance [5]. The wavelength of the micro laser is 2.8 μm. Light with this wavelength is mostly absorbed by water, and therefore this laser is effective only on the surface of brain tissue, enabling precise ablation at the boundary between tumor and normal tissue [4].

In this paper, we developed a combined system of tumor detection using 5-ALA and precise ablation by micro laser, with an automatic focusing (AF) and robotic scanning mechanism for the brain surface. This system is designed for possible localized pinpoint detection of the tumor, then ablating the fluorescent area with stepping driven precise position control in the whole system. This is first attempt to integrate intraoperative fluorescence detection and high precision laser ablation system. In addition, each of the measurement and ablation is performed under the robotic position controlling.

2 Materials and Methods

In this chapter, the tumor detection using 5-ALA-induced PpIX fluorescence, the micro laser module, the automatic focusing and robotic scanning system, and the whole system integration are described. Finally we proposed the experimental procedure for combining these surgical processes.

2.1 Intraoperative Detection System for Brain Tumor Using 5-ALA

Fluorescence of PpIX is fully observed a few hours after orally administrated 5-ALA. Irradiating near-ultraviolet light of around 400 nm, PpIX emits the fluorescence of the wavelength of 635 nm at peak intensity from the brain tissue.

Applying a highly sensitive camera is one of the reasonable ways for intraoperative detection of PpIX fluorescence. Although it is possible to acquire a wide-area 2-dimensional data at one time, obtained data contains only light intensity and simple color information. As PpIX fluorescence is so much weaker than the excitation light and a guide laser of the AF system, it requires the optical narrow-band-pass filter to cut off the other light sources. Nevertheless, it is difficult to separate the fluorescent component from the intensity and the color information of the image signal. Therefore, we chose the use of a spectral photometer, easily extracting the peak intensity of the fluorescence. The spectral data contains important information about tumors; for example, a spectral shape and a peak wavelength possibly vary with a density of tumor cells, a class of tumors, and other conditions of tissues. In the future task, we consider acquisitions of the functional information of tumors by spectral analysis.

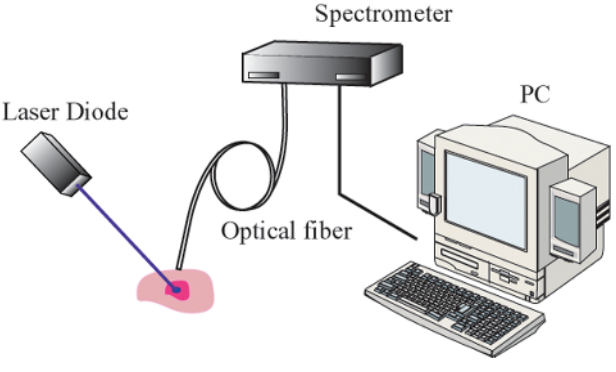


Fig. 1. Tumor detection system using 5-ALA-induces PpIX fluorescence

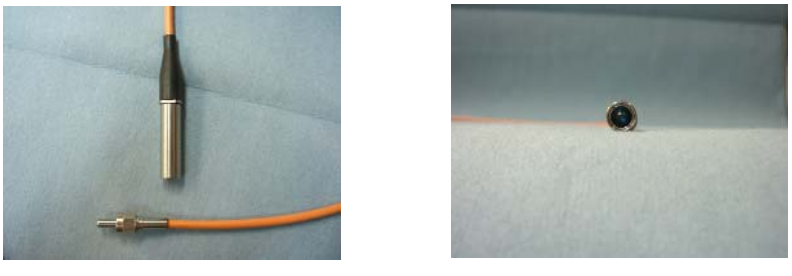


Fig. 2. Fluorescence detection probe

PpIX fluorescence was collected by a detection probe and guided into a spectrophotometer through an optical single-mode fiber, then performed spectral analysis (Fig. 1). The detector has a diameter of 8 mm, using aspheric lenses to correct aberrations. The detection resolution is set to 0.6 mm, considering a tradeoff between not decreasing of the acquired light intensity in proportion to the square of the resolution diameter and improving the accuracy of the measurement (Fig.2). A band-pass filter, which transmits over 60 % at 635 nm and up to 5 % at 670 nm, was

fixed on the tip of the detector, cutting off the excitation light and the guide laser (peak at 670 nm) of the AF system. The working distance of the detector is 16 mm to a tolerance of plus or minus 0.5 mm.

2.2 Micro Ablation Laser Module

For ablation of tumor tissues, we used a mid-infrared continuous-wave laser with a wavelength of 2.8 μm, being output by a microchip solid-state laser on the tip of a laser probe [4]. The pumping light source for the solid-state laser is a near-infrared diode laser with a wavelength of 970 nm, guided through a quartz optical fiber to the laser probe.

As the light wavelength around 3μm has strong absorption feature by water, this laser is effective only on the surface of the living tissue, and it can make a precise ablation with a low output of 0.2 W or less. The laser beam is focused to a diameter of 0.1 to 0.15 mm with a lens, and an ablation groove is formed equivalent to the spot diameter in the soft vital tissue. The working distance of the laser probe is 15 mm ± 1 mm.

2.3 Automatic Focusing and Robotic Scanning System

Both the fluorescence detection probe and the micro laser probe have each working distance, and this requires an AF mechanism, constantly maintaining the distance from the brain surface. In this practice, we used an AF system designed based on the three-dimensional measurement system (Mitaka Ko-ki Co., Ltd.) (Fig. 3). In this system, position measurement was performed using a confocal optical mechanism and the guide laser was picked up with a split photodiode, enabling a focusing with an accuracy of micrometers. The wavelength of the guide laser is 670 nm. This system was coupled with 2-axial automatic stepping drive stage and can make a robotic scanning on the surface of the brain.

Fig. 3. Automatic focusing and robotic scanning system

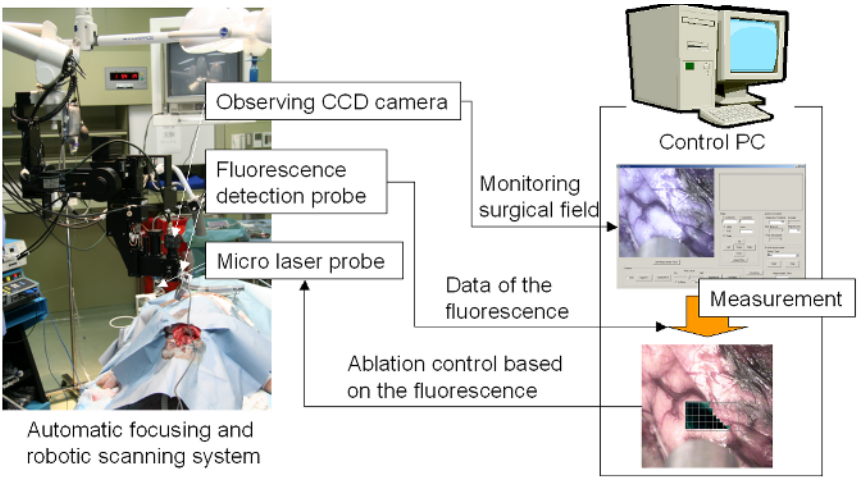


Fig. 4. System configuration

2.4 System Integration

Fig. 4 shows the system configuration in this paper. The fluorescence detection probe and the micro laser probe were attached to the AF system. The data from the spectral photometer was stored into a personal computer (PC). Switching of the micro laser and scanning with the stepping drive were both controlled by the PC. We can observe surgical field view by a CCD camera in the AF system. This image was used to control electric motors to position the fluorescence measurement system and the laser ablation system.

2.5 Experimental Procedure

Measurement of Fluorescence Signal

Measurement area was defined on a CCD camera view of the AF system and sectioned in a grid pattern. Measurement was performed on each grip point with raster

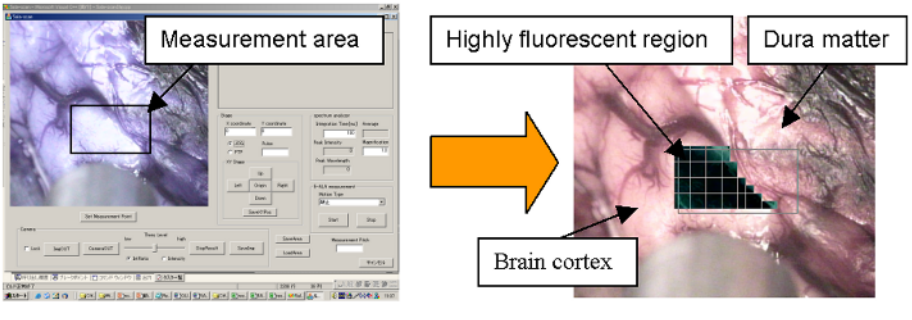


Fig. 5. An example of the measurement procedure of the fluorescence on a porcine brain

scanning. A point in the measurement area was regarded to have same fluorescent property with nearest grid point. A grid interval was determined to 0.4 mm, considering that the grid square was included in the circle of the measurement resolution of 0.6 mm. As a measurement required more than a few hundred of milliseconds, it took several tens of seconds or more to measure a few millimeters square. Fig. 5 shows an example of the fluorescence measurement procedure.

Ablation of the Target Area

A threshold was set for the intensity level of measured fluorescence. In this paper, a surgeon or an experimenter determined the threshold, viewing the measurement area on the CCD camera image. Then scanning was started, and the micro laser was automatically irradiated on the region over the threshold.

3 Results

Combination tests of the tumor detection system and the micro ablation laser module were performed for a biomedical simulant material (phantom) and a porcine brain.

3.1 Phantom Experiment

The phantom was composed of agar plate containing Intralipid-10% that is intravenous lipid emulsion and used for scattering medium [6]. The concentration of Intralipid-10% was adjusted so that the scattering coefficient became 3 cm^{-1}, which is similar to that of Glioma [6][7]. A half part of the phantom contained PpIX to emit fluorescence and the other half part did not contain PpIX.

Fig. 6(a) shows one of the results of experiments. Black rectangular area stands for the scanned area by the system. The system could identify the boundary between the area with fluorescence and without fluorescence, and could precisely ablate the fluorescent area with automatically scanning.

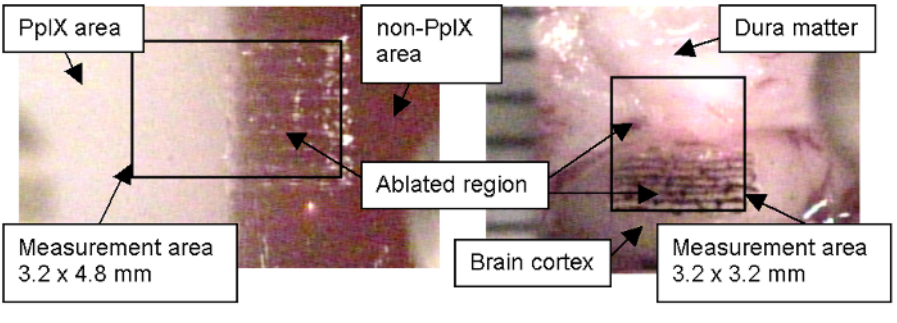


Fig. 6. Result of the combination tests of the tumor detection system and the micro ablation laser module

3.2 Porcine Brain Experiment

The target was the surface of a porcine brain exposed by craniotomy under anesthesia. Before the experiment, 5-ALA was administrated in sufficient quantities to accumulate on a normal brain tissue and metabolize to become PpIX. A half of the measurement area was covered with dura matter, where the fluorescence was not observed, and the other half was exposing brain cortex.

Fig. 6(b) shows one of the results of experiments. Black rectangular area stands for the scanned area by the system. The dura matter covering the porcine brain could not completely block the fluorescence. We could identify weak signal of fluorescence from some part of the dura matter. As shown in Fig. 6(b), ablation laser was irradiated to such area on the covered region. We could successfully identify and ablate the fluorescent area of the porcine brain based on the fluorescence data. The AF system functioned properly even for porcine brain and was effective to stabilize both of the conditions of the measurement and the laser ablation.

4 Discussion and Conclusions

We developed a combined system of tumor detection by 5-ALA-induced PpIX fluorescence and precise ablation by micro laser for the first time, with an automatic focusing and robotic scanning mechanism for the brain surface. In this system, ablation was performed based on the fluorescent information under the robotic position controlling.

Combination tests of the fluorescence measurement and the laser ablation were performed for a biomedical simulant material (phantom) and a porcine brain. Measurement areas of the phantom and the porcine brain were both separated into fluorescent part and non-fluorescent part. In each test, it was possible to extract the area with fluorescence appropriately from the measurement data, and the extracted area was selectively ablated by the micro laser with automatically scanning.

In this practice, the experimental targets were clearly separated into the area with fluorescence and without fluorescence. In clinical cases, the boundaries between tumors and normal tissues are often unclear, and tumors invade normal tissues. Therefore, thresholding of a fluorescence data and an extraction of the area for ablation will have problems, and are considered making some automation process. Solutions for these problems are desired in the future work. To automatically discriminate the tumor region by fluorescence data, a multiple classification analysis based on another spectral features not only peak intensity of the fluorescence will be investigated.

Acknowledgments

We wish to thank many people and groups who have contributed to this effort. Research on medical devices for analyzing, supporting and substituting the function of human body funded by Ministry of Health, Labour and Welfare.

References

1. Clatz Oliver, et. al, Robust nonrigid registration to capture brain shift from intraoperative MRI, IEEE Transactions on Medical Imaging, vol. 24, no. 11, November, pp. 1417-1427, 2005.
2. T. Maruyama, et al, "Intraoperative detection of malignant gliomas using 5-Aminolevulinic acid induced protoporphyrin fluorescence, openMRI and real-time navigation system," Computer Assisted Radiology and Surgery, vol. 15, pp. 270-275, 2001.
3. K. Shimizu, et al, "Application of blue semiconductor laser to measurement of 5-ALA induced fluorescence for intraoperative detection of brain tumor," Proceedings of 6th Japan-France Congress on Mechatronics and 4th Asia-Europe Congress on Mechatronics, 2003, pp. 135-140.
4. S. Omori, Y. Muragaki, I. Sakuma, and H. Iseki, "Robotic laser surgery with λ = 2.8um microlaser in neurosurgery," Journal of Robotics and Mechatronics, vol. 16, no. 2, pp. 122-128, 2004.
5. M. Norio, et al, In light of recent developments, application of fluorescence spectral analysis in tumor diagnosis, Applied Spectroscopy Reviews, vol. 39, no. 4, November, pp. 437-455, 2004.
6. J. van Staveren, J. M. Moes, J. van Marle, A. Prahl, J. C. van Germert, "Light scattering in Intralipid-10% in the wavelength range of 400-1100 nm," Appl Opt, vol. 30, pp. 4507-4514, 1991.
7. Sterenborg HJ, et al, "The spectral dependence of the optical properties of the human brain," Lasers Med Sci, vol. 4, pp. 221-227, 1989.

Visual Servoing for Intraoperative Positioning and Repositioning of Mobile C-arms

Nassir Navab[1], Stefan Wiesner[1], Selim Benhimane[1], Ekkehard Euler[2], and Sandro Michael Heining[2]

[1] Chair for Computer Aided Medical Procedures (CAMP), TU Munich, Germany
[2] Chirurgische Klinik und Poliklinik - Innenstadt, LMU Munich, Germany

Abstract. The problem of positioning mobile C-arms, e.g. for down the beam techniques, as well as repositioning during surgical procedures currently requires time, skill and additional radiation. This paper uses a Camera-Augmented Mobile C-arm (CAMC) to speed up the procedure, simplify its execution and reduce the necessary radiation. For positioning the C-arm in down-the-beam position, the pre-operative diagnostic CT is used for defining the axis. Additional CT visible markers on patient's skin allow the CAMC's optical camera to compute the C-arm's pose and its required displacement for positioning. In the absence of electronically controlled mobile C-arms, the system provides step-by-step guidance to surgical staff until the final position is achieved. At this point, the surgeon can acquire an X-ray to ensure the correct positioning. In the case of intra-operative repositioning, no pre-operative CT is required. X-ray/Optical markers allow the visual servoing algorithm to guide the surgical staff in C-arm repositioning using CAMC's optical camera. This work paves the path for many possible applications of visual servoing in C-arm positioning and in surgical navigation. Experiments on phantom and a cadaver study demonstrate the advantages of the new methods.

1 Introduction

The mobile C-arm is an essential everyday tool in routine trauma and orthopaedic surgery. With increasing numbers of minimally invasive procedures the importance of fluoroscopic guidance is still growing, thus augmenting radiation doses [1,2,3,4]. Considerable effort has been undertaken to improve the possibilities of C-arm imaging. Nowadays intraoperative 3D-imaging and C-arm based surgical navigation systems are commercially available. However in the surgical workflow, 3D-imaging is only possible at distinct points during the intervention e.g. to visualize the quality of fracture-reduction or to control the position of implants. During 3D image acquisition no manipulation like drilling or implant positioning is possible, i.e. the different steps of the surgical procedure are still carried out under 2D fluoroscopic imaging. So radiation exposure both to the patient and the surgical staff is often inevitable. In some surgical procedures even the direct exposure of the surgeon's hand cannot be avoided.

In the last decade the first medical augmented reality (AR) systems were introduced [5,6,7,8,9,10]. Most of the systems augment the view of the surgeon

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 551–560, 2006.

or an external camera with registered pre-operative data. Four Medical AR systems were also proposed that directly augment the intra-operatively acquired data [11,12,13,14]. Camera-Augmented Mobile system is one of these systems, which attaches an optical camera to the X-ray source housing. Thanks to a double mirror system and an adopted one time calibration procedure performed during the construction of the system, the X-ray and optical images are aligned for all simultaneous acquisitions. If the patient does not move, the video images remain aligned with the X-ray image. This makes the system quite interesting for medical applications. We have built a new CAMC system based on the description provided in [11,15]. We have also demonstrated the efficiency of this video augmented X-ray fluoroscopy system for pedicale screw placement through a cadaver study [16]. This work aims at the use of this newly built system for positioning and repositioning of the mobile C-arms during surgical procedures. The main contribution here is the use of visual servoing with no radiation for positioning a mobile C-arm to obtain target images defined for the X-ray imaging. We have focused on the two following tasks that are often performed in various surgical applications:

Down-the-beam positioning. This task is often performed when there is a need for inserting a linear surgical tool to reach a deep seated anatomical target or to position a screw into an interlocking hole. It consists of aligning the entry point and a deep seated target, or the axis of perforation with the X-ray source such that the linear surgical tool or the axis of perforation is projected into a single point or the interlocking hole is presented with maximum apparent diameter on the X-ray image. Clinical examples for this "down-the-beam"-technique include, but are not limited to, the distal interlocking of intramedullary implants in the treatment of long bone fractures [17], the placement of transpedicular screws in spinal surgery [18,16] and precise needle placement.

Intraoperative repositioning of mobile C-arm. During the course of an intervention, the mobile C-arm has to be moved by OR-staff. These movements include bringing the C-arm into the operation field and out again at certain points of the procedure, when fluoroscopic control is needed as well as the positioning for specific projections in regard to the anatomy, e.g. joint lines. For the control of the correct positioning, supporting systems such as goniometers or laser-visors are used. However additional fluoroscopic images cannot be avoided.

In this paper, we propose two visual servoing algorithms providing elegant solutions for performing the above two tasks with no additional radiation for surgeon as well as for the patient. Experimental results prove the efficiency of the algorithms.

2 Overview of Methods

In general, we attack the problem of finding a predefined position for the mobile C-arm with respect to the patient. Markers are attached to the patient, which will be used as features for the visual servoing tasks.

The repositioning task is done by (optical) image-based visual servoing. A target image from the current C-arm position is taken. When the C-arm is repositioned, the current positions of the fiducials are compared to the target positions and joint increments of the C-arm are computed (see figure 1).

For Down-the-Beam positioning of the C-arm, we use a pre-operative CT to define the Down-the-Beam axis. Intra-operatively we estimate the current pose of the C-arm with regards to the CT data (see figure 2). Position-based visual servoing is used to move the C-arm towards the desired position. When we are close to the optimal position, we switch to image-based visual servoing with the target image generated from the CT data .

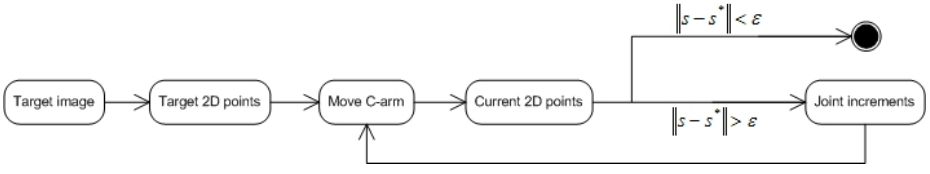


Fig. 1. intra-operative repositioning

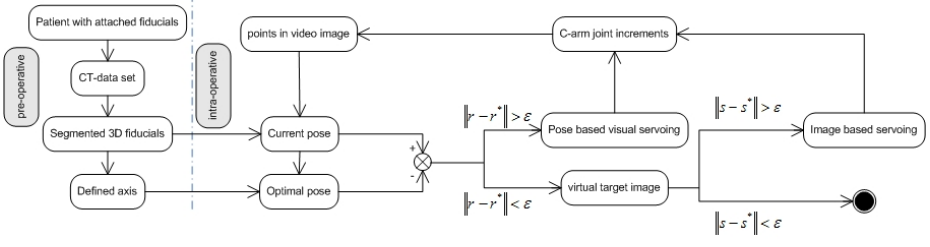


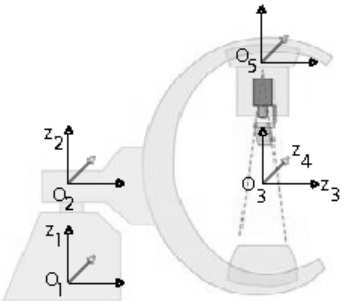
Fig. 2. Down-the-Beam C-arm positioning

3 Visual Servoing

Visual servoing aims to control the C-arm using the visual information taken from the on-board camera observing the scene at the current position $\mathcal{F}$ such that it reaches the desired (or the reference) position $\mathcal{F}^*$. If we denote by $\mathbf{q}$ the vector containing the joints current position, the objective of visual servoing is then to compute the direction and the amplitude of the increments of the joints, that correspond to $\dot{\mathbf{q}}$, in order to accomplish the positioning task. Consequently, we need to model the forward kinematics and the Jacobian of the C-arm.

3.1 C-arm Motion Model

In Figure 3, we sketched the forward kinematics of the C-arm (a similar kinematic analysis has also been done in [19]). The C-arm has 5 degrees of freedom:



Tz	Rz	Tx	Rx
0	wigwag	0	0
height	0	0	-pi/2
length+offset	-pi/2	0	0
0	angular	0	pi/2
0	orbital	offset	0

Fig. 3. C-arm coordinate frames

the height, the wigwag-movement, the length, the angular and the orbital movement. We assigned a coordinate frame to each joint according to the Denavit-Hartenberg rules [20]. Let ${}^{i}\mathbf{A}_{i+1}$ be the (4×4) transformation matrix from the coordinate system of the joint $i+1$ to the coordinate system of the joint i. Multiplying the different transformation matrices makes it possible to obtain the camera pose wrt. the world reference frame: ${}^{0}\mathbf{A}_{6} = {}^{0}\mathbf{A}_{1}{}^{1}\mathbf{A}_{2}{}^{2}\mathbf{A}_{3}{}^{3}\mathbf{A}_{4}{}^{4}\mathbf{A}_{5}{}^{5}\mathbf{A}_{6}$ The transformation matrix ${}^{5}\mathbf{A}_{6}$ takes into account the rotation between the coordinate system of the last joint and the coordinate system of the camera.

To relate the motion of the camera to the motion of the C-arm joints, we have to compute the manipulator Jacobian $\mathbf{J}_{carm}$ of the C-arm. It converts the velocities of the single joints $\dot{\mathbf{q}}$ to the Cartesian velocity of the camera $\mathbf{v}_c$.

The C-arm consists of 5 joints so the manipulator Jacobian is a 6x5 matrix. Its entries can be derived from the forward kinematics. The rotation axis $\mathbf{z}_i$ of each joint is found in the 3rd column of the matrix ${}^{0}\mathbf{A}_{i} = {}^{0}\mathbf{A}_{1} \ldots {}^{i-1}\mathbf{A}_{i}$, where the origin $\mathbf{o}_i$ is found in the 4th column. The axis and origins are visualized in Figure 3. Equation 1 lists first the column of the manipulator Jacobian for a prismatic joint (like length and height of the C-arm) and second for a revolute joint (like wigwag, angular and orbital movement).

$$\mathbf{j}_i = \begin{pmatrix} \mathbf{z_i} \\ \mathbf{0} \end{pmatrix} \; ; \mathbf{j}_i = \begin{pmatrix} \mathbf{z_i} \times (\mathbf{o}_5 - \mathbf{o}_i) \\ \mathbf{z_i} \end{pmatrix} \tag{1}$$

3.2 Camera Model

At the desired position $\mathcal{F}^*$, a 3D point $\boldsymbol{\mathcal{X}}$ is projected on a virtual plane perpendicular to the optical axis of the camera into a 2D point $\mathbf{m}^*$:

$$\mathbf{m}^* = (x^*, y^*, 1) \propto \left[\mathbf{I}_{3\times3}\ \mathbf{0}_{3\times1}\right] \boldsymbol{\mathcal{X}} \tag{2}$$

The same 3D point $\boldsymbol{\mathcal{X}}$ is projected into a 2D point $\mathbf{m}$ in the current camera position $\mathcal{F}$: $\mathbf{m} = (x, y, 1) \propto \left[\mathbf{R}\,\mathbf{t}\right] \boldsymbol{\mathcal{X}}$ where $\mathbf{R}$ is the rotation matrix and $\mathbf{t}$ is the translation vector between the two coordinate systems $\mathcal{F}$ and $\mathcal{F}^*$. The information given by a pinhole camera, which performs a perspective projection of 3D points, is an image point $\mathbf{p} = (u, v, 1)$ verifying $\mathbf{p} = \mathbf{K}\mathbf{m}$ where $\mathbf{K}$ is the camera internal parameter matrix.

3.3 Pure Visual Servoing Using 2D Target Image

The pure visual servoing using 2D target image can be done by making a vector $\mathbf{s}$, containing visual information extracted from the current acquired image (at the current position $\mathcal{F}$), converging to a vector $\mathbf{s}^*$ containing visual information extracted from the reference image (at the reference position $\mathcal{F}^*$). In our case, the visual information are the image coordinates of the n markers:

$$\mathbf{s} = \begin{bmatrix} \mathbf{m}_1^\top & \mathbf{m}_2^\top & \dots & \mathbf{m}_n^\top \end{bmatrix}^\top \tag{3}$$

An interaction matrix $\mathbf{L}$ is then defined in order to establish the relationship between the Cartesian velocity of the camera $\mathbf{v}_c$ and the derivative of the vector $\mathbf{s}$ wrt. time. This relationship can be written as: $\dot{\mathbf{s}} = \mathbf{L}\mathbf{v}_c$. The $(3n \times 6)$ interaction matrix $\mathbf{L}$ can be expressed with the following formula: $\mathbf{L} = \begin{bmatrix} \mathbf{L}_1^\top & \mathbf{L}_2^\top & \dots & \mathbf{L}_n^\top \end{bmatrix}^\top$ where:

$$\mathbf{L}_i = \begin{bmatrix} \frac{1}{Z_i} & 0 & -\frac{x_i}{Z_i} & x_i y_i & -(1+x_i^2) & y_i \\ 0 & \frac{1}{Z_i} & -\frac{y_i}{Z_i} & (1+y_i^2) & -x_i y_i & -x_i \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The vector $\mathbf{s}$ is servoed using the task function approach [21,22] by minimizing iteratively the following vector $\mathbf{e} = \widehat{\mathbf{L}}^+(\mathbf{s} - \mathbf{s}^*)$, where $\widehat{\mathbf{L}}^+$ is an approximation of the pseudo-inverse of the true interaction matrix $\mathbf{L}^+ = (\mathbf{L}^\top \mathbf{L})^{-1}\mathbf{L}^\top$. If we differentiate this equation, we obtain:

$$\dot{\mathbf{e}} = \frac{d\widehat{\mathbf{L}}^+}{dt}(\mathbf{s} - \mathbf{s}^*) + \widehat{\mathbf{L}}^+ \dot{\mathbf{s}} = (\mathbf{O}(\mathbf{s} - \mathbf{s}^*) + \widehat{\mathbf{L}}^+ \mathbf{L})\mathbf{v}_c \tag{4}$$

where $\mathbf{O}(\mathbf{s} - \mathbf{s}^*)$ is a (6×6) matrix that can be neglected for $\mathbf{s} \approx \mathbf{s}^*$. Let's consider the control law: $\mathbf{v}_c = -\lambda \mathbf{e}$ where λ is a positive scalar. Plugging this equation into equation (4), we obtain the following closed-loop equation: $\dot{\mathbf{e}} = -\lambda(\mathbf{O}(\mathbf{s} - \mathbf{s}^*) + \widehat{\mathbf{L}}^+ \mathbf{L})\mathbf{e}$. It is well known from the control theory that this non-linear system is locally asymptotically stable in a neighborhood of $\mathbf{s} = \mathbf{s}^*$, if and only if, the matrix $\widehat{\mathbf{L}}^+ \mathbf{L}$ has eigenvalues with a positive real part $real(eig(\widehat{\mathbf{L}}^+ \mathbf{L})) > 0$. Obviously, if the pseudo-inverse of the interaction matrix is well approximated, we have: $\widehat{\mathbf{L}}^+ \mathbf{L} \approx \mathbf{I}$ and the control law is locally asymptotically stable. Then, using the camera Cartesian velocity $\mathbf{v}_c$, it is possible to compute the joints increments $\dot{\mathbf{q}}$ given the pseudo-inverse of the C-arm Jacobian:

$$\dot{\mathbf{q}} = \mathbf{J}_{carm}^+ \mathbf{v}_c \tag{5}$$

3.4 Position-Based Visual servoing

Figure 4.a shows the algorithm to find the optimal C-arm pose. First the initial C-arm pose ${}^w\mathbf{T}_c$ is set manually by inputting a rough estimate of the initial joint values, since no sensors are attached to the C-arm, which report the current pose. Then we estimate the current pose ${}^c\mathbf{T}_{ct}$ of the camera wrt. to the frame of the CT-data set. The data is now transformed into the world frame $\boldsymbol{\mathcal{X}}_w = {}^w\mathbf{T}_c$

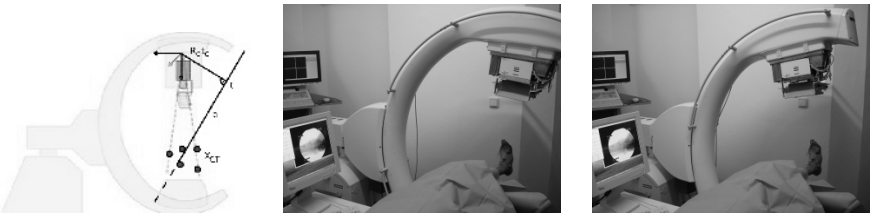


Fig. 4. (a) Projection of current camera center onto defined axis (b) reference position and start position of the C-arm

${}^{c}\mathbf{T}_{ct}\boldsymbol{\mathcal{X}}_{ct}$ To get a first approximation of the optimal camera position, the camera center $\mathbf{t}$ is projected onto the defined axis $\mathbf{a}$. This gives the new camera center $\mathbf{t}^*$. The translational velocity is then set such that it is proportional to the difference between the current and the desired coordinate system center. While the rotational velocity $\boldsymbol{\omega}$ is computed thanks to the cross-product between the current camera optical axis $\mathbf{z}$ and the axis $\mathbf{a}$: $\boldsymbol{\nu} = -\lambda_\nu(\mathbf{t} - \mathbf{t^*})$ and $\boldsymbol{\omega} = -\lambda_\omega(\mathbf{a} \times \mathbf{z})$ We now assemble them into the velocity screw. The joint increments are computed by equation 5. Now the C-arm is moved according to the computed parameters and the optimal position is estimated again. We stop the iterations when we are near the optimal position and generate a target image which is used for the final adjustment based on image based visual servoing.

3.5 Generation of the Target Image

In order to generate a virtual target image, we set up a virtual camera at the position $\mathbf{T}^*$: with $\mathbf{a}$ the direction of the segmented axis, $\mathbf{u}$ and $\mathbf{v}$ verify $\mathbf{u} \times \mathbf{v} = \mathbf{a}$ and $\mathbf{p}$ is a point on the axis. The remaining parameters, a rotation θ around and a translation t along the segmented axis are encapsulated in the transformation matrix $\mathbf{T}(\theta, t)$

$$\mathbf{T}^* = \begin{bmatrix} \mathbf{u}\ \mathbf{v}\ \mathbf{a}\ \mathbf{p} \\ 0\ 0\ 0\ 1 \end{bmatrix} \text{ and } \mathbf{T}(\theta, t) = \begin{bmatrix} cos(\theta) & -sin(\theta) & 0 & 0 \\ sin(\theta) & cos(\theta) & 0 & 0 \\ 0 & 0 & 1 & t \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{6}$$

The parameters are chosen such that they minimize the cost function $\min_{\theta,t} \|\mathbf{s} - \mathbf{s}^*(\theta, t)\|$ where $\mathbf{s}$ contains the current image points (given by the equation (3)) and $\mathbf{s}^*(\theta, t)$ contains the projection of the computed 3D points

$$\mathbf{s}^*(\theta, t) = \left[\, \mathbf{m}_1^{*\top}(\theta, t)\ \mathbf{m}_2^{*\top}(\theta, t) \ldots\ \mathbf{m}_n^{*\top}(\theta, t) \,\right]^\top \tag{7}$$

with $\forall i$, $\mathbf{m}_i^*(\theta, t)$ is the projection of the i-th 3D point given by the CT-data set and transformed by $\mathbf{T}^*\mathbf{T}(\theta, t)$.

4 Experiments

4.1 Down-the-Beam Positioning

A spine phantom was placed under the C-arm in neutral position. The pose is estimated and the video image is augmented with the preoperative segmented data (fiducials and axis). Following the instruction of the servoing algorithm we iteratively changed the joint parameters. Figure 5(1-4) shows the acquired images during position based guidance. We performed first an orbital movement, a length change, an angular movement and finally again a length change. Being close to the optimal C-arm position we generated the target image and superimposed it onto the video image. Fine adjustment brings the C-arm to the final position, where an X-ray is taken and blended into the view. Now the placement of the screw is performed down the beam (see figure 5).

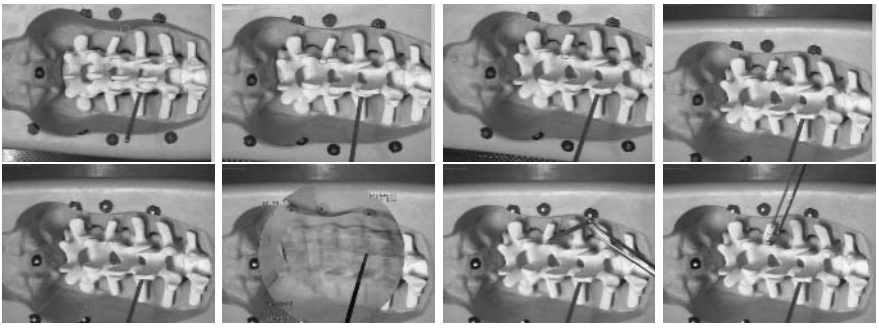


Fig. 5. position based servoing; 1: start image; 2: orbital and length change; 3: angulation change; 4:length change 5: target image superimposition; 6: length changed and X-ray superimposition; 7,8: placement of pedicle screw

4.2 C-arm Repositioning: Cadaver Study with Ankle Joint

The standard radiographic examination of the ankle for fracture detection consists of the ap, mortise and lateral view. Intraoperatively the precise mortise view is mandatory to proof the correct reposition of the ankle. This view can be achieved at approximately 20° internal rotation. The C-arm was positioned by the surgeon according to the intervention (ap,mortise). After that we moved the C-arm position about $-20°$ in the orbital direction and $-5°$ in the angular one (figure 4.b). To reposition the C-arm we followed the instruction of the visual servoing algorithm. Fine adjustment was done by comparing the current fiducial position with the target ones. C-arm reposition was performed according to the computed instructions: start with orbital movement, followed by a wigwag motion. The next iteration computed an angular movement and a wigwag movement. The reference X-ray and the resulting X-ray acquired after repositioning are fused with a checkerboard blending (Figure 7) to verify the result.

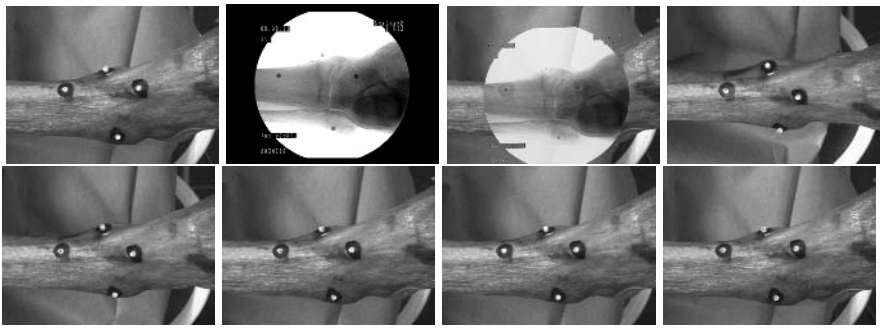


Fig. 6. Ankle joint repositioning; 1: reference video; 2: reference X-ray; 3: blended reference images; 4-8: images acquired during movement from start to reference position of C-arm

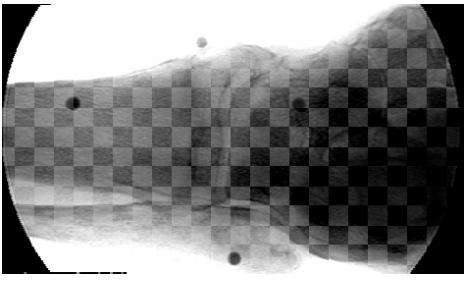


Fig. 7. Checkerboard fusion of reference X-ray and repositioning X-ray

5 Discussion

The presented positioning algorithm proved to be robust and effective in a clinical setup (pedicle screw placement and ORIF of ankle fracture). Computation is fast (real-time) and thus easily acceptable for clinical applications. After guidance with help of the algorithm interactive visual servoing allows high precision at small incremental steps. The number of positioning steps is comparable to clinical routine but without any radiation exposure. However X-ray images can be obtained at any time to check the reached position or to check patient movement. On one hand, the system fixes user's errors and takes patient's movements into account by systematically recomputing the appropriate trajectory at each step. On the other hand since the surgeon is also in the loop, rare potential risks, such as collision with the OR table, are prevented. The solution is intuitive and can be easy integrated into the clinical workflow.

6 Conclusion

We proposed a visual servoing approach for positioning and repositioning of the mobile X-ray C-arm with attached and co-registered optical camera. The C-arm geometry and all its degrees of freedom are modelled and two solutions for position based and image based visual servoing are proposed. We have tested the system in collaboration with a trauma surgeon in two experiments, the first one on a phantom and the second one on a cadaver. Interestingly, the surgeon prefers the interactive solution more than a possible full automatic one, since the process is quite fast and he is in full control. The system allows frequent tasks such as C-arm repositioning and down-the-beam positioning to be performed with minimum radiation for patient and no radiation for physician.

References

1. Boszczyk, B.M., et al.: Fluoroscopic radiation exposure of the kyphoplasty patient. European Spine Journal (2006) 347 – 355
2. Synowitz, M., Kiwit, J.: Surgeon's radiation exposure during percutaneous vertebroplasty. J Neurosurg Spine. **4**(2) (2006) 106–109
3. Rampersaud, Y.R., et al.: Radiation exposure to the spine surgeon during fluoroscopically assisted pedicle screw insertion. Spine. **25**(20) (2000) 2637–2645
4. Theocharopoulos, N., et al.: Occupational exposure from common fluoroscopic projections used in orthopaedic surgery. The Journal of Bone and Joint Surgery (American) **85** (2003) 1698–1703
5. State, A., Chen, D.T., Tector, C., Brandt, A., Chen, H., Ohbuchi, R., Bajura, M., Fuchs, H.: Case study: observing a volume rendered fetus within a pregnant patient. In: Proceedings of the conference on Visualization '94, IEEE Computer Society Press (1994) 364–368
6. Grimson, W., Ettinger, G., Kapur, T., M.E. Leventon, W.M. Wells, R.K.: Utilizing segmented mri data in image guided surgery. International Journal of Pattern Recognition and Artificial Intelligence (1996)
7. P.J.Edwards, Hill, D.D., Hawkes, D.D., Colchester, D.A.: Neurosurgical guidance using the stereo microscope. In: Proceedings of Computer Vision, Virtual Reality and Robotics in Medicine, First International Conference, CVRMed'95. (1995)
8. Blackwell, M., Nikou, C., Gioia, A.M.D., Kanade, T.: An image overlay system for medical data visualization. (1998)
9. Sauer, F., Schoepf, U.J., Khamene, A., Vogt, S., Das, M., Silverman, S.G.: Augmented reality system for ct-guided interventions: System description and initial phantom trials. In: Medical Imaging: Visualization, Image-Guided Procedures, and Display. (2003)
10. Birkfellner, W., Figl, M., Huber, K., Watzinger, F., Wanschitz, F., Hummel, J., Hanel, R., Greimel, W., Homolka, P., Ewers, R., Bergmann, H.: A head-mounted operating binocular for augmented reality visualization in medicinedesign and initial evaluation. IEEE Transaction on Medical Imaging **21**(8) (2002) 991–997
11. Navab, N., Mitschke, M., Bani-Hashemi, A.: Merging visible and invisible: Two camera-augmented mobile C-arm (CAMC) applications. In: Proc. IEEE International Workshop on Augmented Reality, San Francisco, CA, USA (1999) 134–141
12. Stetten, G.D., et al.: C-mode real time tomographic reflection for a matrix array ultrasound sonic flashlight. (2003)

13. Masamune, K., et al.: An image overlay system with enhanced reality for percutaneous therapy performed inside ct scanner. (2002)
14. Leven, J., et al.: DaVinci Canvas: A Telerobotic Surgical System with Integrated, Robot-Assisted, Laparoscopic Ultrasound Capability. In: MICCAI. (2005) 811–818
15. Mitschke, M., Bani-Hashemi, A., Navab, N.: Interventions under video-augmented X-ray guidance: Application to needle placement, Pittsburgh, PA, USA (2000) 858–868
16. Heining, S., Euler, E., Wiesner, S., Navab, N.: Pedicle screw placement under video-augmented fluoroscopic control: First clinical application - a cadaver study. In: CARS 2006 - Computer Assisted Radiology and Surgery. (2006)
17. Whatling, G.M., Nokes, L.D.: Literature review of current techniques for the insertion of distal screws into intramedullary locking nails. Injury. (2005)
18. Hart, R.A., et al.: Pedicle screw placement in the thoracic spine: a comparison of image-guided and manual techniques in cadavers. Spine. **30**(12) (2005) 326–331
19. Binder, N., Matthus, L., Burgkart, R., Schweikard, A.: A robotic c-arm fluoroscope. Int. Journal on Medical Robotics and Computer Assisted Surgery **1**(3) (2005) 108–116
20. Denavit, J., Hartenberg, R.S.: A kinematic notation for lower-pair mechanisms based on matrices. Jour. of Applied Mechanics **22** (1955) 215–221
21. Samson, C., Le Borgne, M., Espiau, B.: Robot Control: the Task Function Approach. Oxford Engineering Science Series. Clarendon Press, Oxford, UK (1991)
22. Espiau, B., Chaumette, F., Rives, P.: A new approach to visual servoing in robotics. IEEE Trans. on Robotics and Automation **8**(3) (1992) 313–326

Navigated Three Dimensional Beta Probe for Optimal Cancer Resection

Thomas Wendler[1], Joerg Traub[1], Sibylle Ilse Ziegler[2], and Nassir Navab[1]

[1] Chair of Computer Aided Medical Procedures (CAMP), TU Munich, Germany
twendler@ieee.org, {traub, navab}@cs.tum.edu
[2] Nuclear Medicine Department, Klinikum rechts der Isar der Technische Universität München, Germany
s-ziegler@lrz-muenchen.de

Abstract. In minimally invasive tumor resection, the desirable goal is to perform a minimal but complete removal of cancerous cells. In the last decades interventional nuclear medicine probes supported the detection of remaining tumor cells. However, scanning the patient with an intraoperative probe and applying the treatment are not done simultaneously. The main contribution of this work is to extend the one dimensional signal of a beta-probe to a four dimensional signal including the spatial information of the distal end of the probe. We generate a color encoded surface map of the scanned activity and guide any tracked surgical instrument back to the regions with measured activity. For navigation, we implement an augmented reality visualization that superimposes the acquired surface on a visual image of the real anatomy. Alternatively, a simulated beta-probe count rate in the tip of a tracked therapeutic instrument is simulated showing the count number and coding it as an acoustic signal. Preliminary tests were performed showing the feasibility of the new designed system and the viability of such a three dimensional intraoperative molecular imaging modality.

1 Introduction

A challenging task in tumor resection is the minimal but complete removal of malignant cells. In the ideal case, the surgeon performs an extirpation such that the security margins are maximized to avoid recurrence, while minimizing the impact on healthy tissue [1, 2].

The standard procedure to confirm a complete removal is a pathological examination of the extracted tissue during the intervention. If the pathological examination detects cancerous cells at the margins, in the absence of spatial information, the surgeon can do nothing but enlarge the overall resection area.

Recently beta-probes have been introduced to improve the detection of remanent malignant tumor cells by means of nuclear labeling [3, 4]. A cancer marker like FDG[1], is injected into the patient before surgery. After the standard resection the surgeon scans the margins with the beta-probe. The detection of

[1] ^{18}F-labeled 2-deoxy-2-fluoro-d-glucose, a positron-emitting glucose analogue and standard tracer for malignancy localization and staging [1].

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 561–569, 2006.

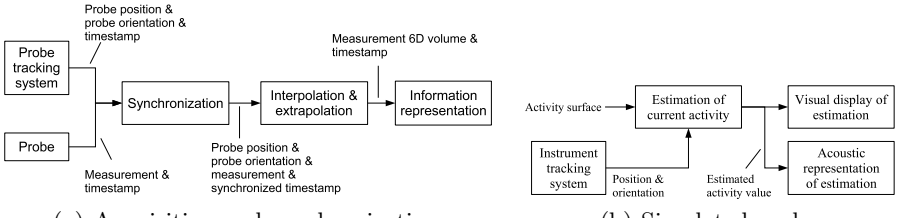
unexpected high activity levels shows the occurrence of residual tumor tissue[2]. The physician has to memorize then the afflicted region, take away the probe, and apply the treatment with a surgical instrument at the place where the highest activity was measured.

The main contribution of this work is to improve the interventional use of beta-probes by combining it with a position and orientation tracking system to generate an activity encoded surface, and to use advanced visualization to represent the functional information to the surgeon. Thus, this work extends the one dimensional beta-probe device to a three dimensional imaging modality. In this paper we explain a prototypical system for recording the measured activity with its corresponding position, and either visualize it by superimposing the generated surface on an image of the real world or by simulating the beta-probe count rate at the tip of any tracked surgical instrument.

2 Requirements and Design Considerations

Hand held probes are commonly used as intraoperative diagnostic devices. They are capable of scanning arbitrary surfaces with a spatial accuracy only limited by the sensor size. Beta-probes are a specially interesting variant, since positrons (beta-plus-particles) penetrate tissue only up to 2 $[mm]$ [3, 5]. This justifies the assumption that the activity measured is located at the surface next to the distal end of the probe and thus enables good spatial accuracy in detection.

Using a position localization system, e.g. an optical tracking system, we can acquire and synchronize recordings of spatial information and beta-particle counts. The combination of both data streams results in a three dimensional imaging device for arbitrary surfaces, which precisely memorizes the synchronized information of activity distribution, and position of the probe on the surface of the examined anatomy. A block diagram of a general navigated probe is shown in figure 1(a).



(a) Acquisition and synchronization. (b) Simulated probe.

Fig. 1. (a) The acquisition of position and orientation of the measurement device and the radioactive counts is recorded and synchronized, resulting in the generation of the three dimensional surface of the activity. (b) The presentation of the recorded activity at the tip of any tracked instrument is done either by displaying the count rate or using an acoustic signal

[2] Normally, the absolute value is not relevant, if it is high relative to the background.

When designing a tracked beta-probe, we must consider the following issues:

1. Since the input is an asynchronous data flow, a proper synchronization system is required. The output is a synchronized four dimensional signal or even seven dimensional, when considering the orientation of the probe. Since we use the constraint that the probe will be moved only orthogonal to the surface, here we neglect the orientation. Furthermore, the detection of the positrons is rather insensitive to small changes in orientation compared to changes in distance[3].
2. The data acquisition is not done on a uniformly spaced grid for position. Thus the surface map has different levels of details depending on the performed scan. One could use interpolation or extrapolation of the measurement in order to fill the missing data. This is optional, and has to be done with care and be subject to medical evaluation.
3. Beta-probes are not point sensors. They have usually a sensor disc with approximately 5 $[mm]$ radius. This has to be considered for the visualization. It also limits the detection accuracy to the sensor size. Nonetheless, multiple measurements within the sensor disc area can achieve a more precise detection of the radioactivity below the surface.

The information representation of the generated surface of synchronized data could be realized by visualizing the surface on a monitor superimposed on the real image of the surgery. This is especially useful for laparoscopic surgery since there is already a camera observing the region of interest that could be used to visualize the surface spatially aligned with the real anatomy (see figure 3(e)).

An alternative representation is the simulation of a real probe on the tip of a surgical instrument (figure 1(b)). This could be implemented through an acoustic signal like a traditional radioactivity counter, by displaying the counts, or by a combination of both. Thus, for any instrument tip one could simulate the measurement of the probe using the previously acquired data. Since both are tracked in the same coordinate system no additional registration step is required.

3 Implementation of a Three Dimensional Beta-Probe

3.1 Hardware Setup

In our setup (figure 3(a)) we used a beta-probe of IntraMedical Imaging LLC (Los Angeles, USA) with the NodeSeeker control unit. The sensitivity of the device with the used shielding is 4500 [counts per second/MBq] at 1 $[cm]$ of a point-like source of ^{18}F. For tracking, we used an external optical infrared tracking system of ART GmbH (Weilheim, Germany) consisting of four ARTtrack2 cameras. The accuracy in localization for four cameras in a volume of $3 \times 3 \times 3$ $[m^3]$ is 0.4 $[mm]$.

[3] From a geometric analysis, it can be shown that for angles smaller than $\pi/20$ the relative error in the radioactive measurement lies below 15% for the dimensions of the probe used and a radioactivity source 2 $[mm]$ below. This error decreases with increasing distance to the source.

We attached a 5DOF (degrees of freedom) tracking target directly on the axis of the probe as shown in figure 3(b). Instruments, cameras and the phantoms were also provided with infrared markers. The entire software for the navigation was implemented in C++ and OpenGL within our research group.

3.2 Synchronization of Data

Depending on the information representation modality, two or three data streams need to be synchronized: tracking system, radioactivity measurement and (for the augmented reality visualization) the video stream. In order to do so, first the computer for the position tracking system and the NodeSeeker for radioactive counts are synchronized using NTP (network time protocol). The data acquired in the both computers is sent via UDP (user datagram protocol) including the time stamp of the moment of acquisition. In our software, we implemented the concept of a ring buffer and can thus correlate the data with the closest time stamps as implemented in [6]. The radioactive count measurements are acquired during a minimum of 100 $[ms]$ to get a statistically representative signal[4]. Since synchronization is the highest source of error in an augmented reality systems [7], we also synchronize the video and tracking data for a precise visualization. We use again a ring buffer and time stamp correlation.

3.3 Calibration and Registration

The distal end of the probe and tip of the instrument need to be calibrated. For this we acquire various poses of the attached tracking target moving the tip of the probe or instrument around a fixed point in space p_W. This leads to a set of equations $p_W = T_{M \to W}(t)\ p_M$, where $T_{M \to W}(t)$ is the transformation from the target coordinate system M (marker) to the tracking coordinate system W (world) at time t, p_W is a constant point in the tracking coordinate system W, and p_M the same constant point in the target coordinate system M. We obtain thus a linear equation system $Ax = b$, where A is a $3n \times 6$ matrix for n time samples, including the rotation part of $T_{M \to W}$, b is a $3n$ vector including the translation part of $T_{M \to W}$, and x is the vector $(p_t^T\ p_w^T)^T$. Since both, probe and instrument, are tracked in the same coordinate system no further registration step is required to simulate the activity on the tip of the instrument.

For the visual representation, the surface is superimposed on the sequence of a video camera, which needs to be calibrated. For this we attach a 6DOF tracking target to it (see figure 3(b)). With a chessboard pattern, we perform a fully automatic calibration in two steps. Firstly, we estimate the distortion coefficients and intrinsic camera parameters from a set of $n > 10$ frames using a standard algorithm for camera calibration [8]. From this procedure the projection matrix P to map three dimensional points onto the image plane is estimated.

[4] Assuming a Poisson distribution of the radioactive decay, it can be proved that for the estimator $\hat{\lambda} = N/\Delta t$ for the activity, the quotient standard deviation to expected value is $(\sqrt{\lambda}\Delta t)^{-1}$ with λ as real activity, N the number of counts and Δt the observation window.

In the second step, based on the same frames, we compute the rigid transformation $T_{C\to I}$ between the camera target coordinate system C (camera) and one at the camera center I (image). This is done by applying hand eye calibration [9] using corresponding poses of the camera in the chessboard coordinate system and poses of the tracking target in the tracking coordinate system W. Any known point in coordinate system W can thus be projected into the image plane by $p_{2D} = P\, T_{C\to I}\, T_{W\to C}\, p_{3D}$, where p_{2D} is the point on the image plane, p_{3D} the point in the tracking coordinate system W, P the projection matrix, $T_{C\to I}$ the rigid transformation, and $T_{W\to C}$ the transformation from coordinate systems W and C, given by the tracking system.

3.4 Information Representation

Given the type of display, the 4D vector list (i.e. activity and 3D position) can be interpolated for the representation of the information. Depending on the representation, we use two different methods for interpolation/extrapolation.

Visual display using virtual or augmented reality: For the visual display, we use the standard Delaunay triangulation for a fast surface rendering over triangles. The nodes in 3D are the real positions of the beta-probe tip and their values are the real measurements. The color of the surface encodes the level of activity. Inside a triangle, the color is linearly interpolated using the distance from the point to be interpolated to the corners as weighting factors[5]. In case of a virtual reality visualization the generated surface is displayed using OpenGL. For the augmented reality, we use as the background texture the real camera image (e.g. of a tracked laparoscope) as in figure 3(e). The projection parameters in OpenGL are set to be the ones estimated during camera calibration. The camera center in the virtual scene is controlled by the pose data $T_{W\to C}$ provided by the tracking system and the calibrated rigid offset $T_{C\to I}$. The block diagram in figure 2 shows the work flow of the visual display for augmented and virtual reality modalities.

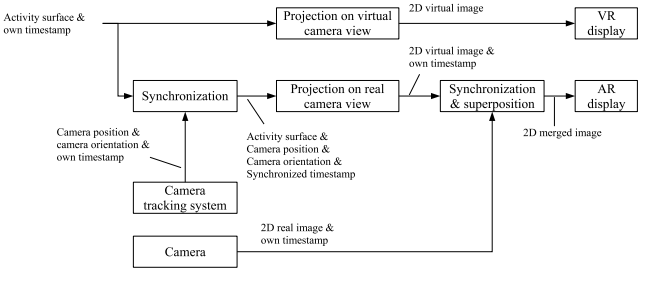
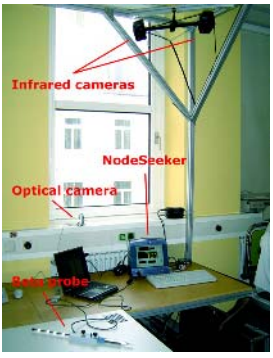


Fig. 2. Block diagram of the visual display using virtual and augmented reality

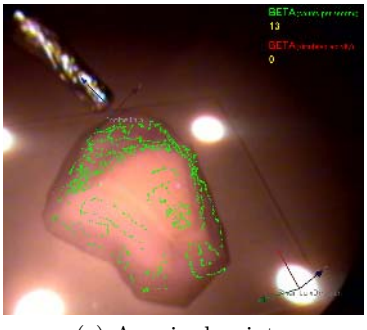
Simulated probe: The second method for representing the acquired information is to simulate the acquired data on the tip of any medical instrument e.g. a knife

[5] One can also visualize the point measurements without interpolation. This interpolation provides uniform activity surfaces, but has to be validated for medical use.

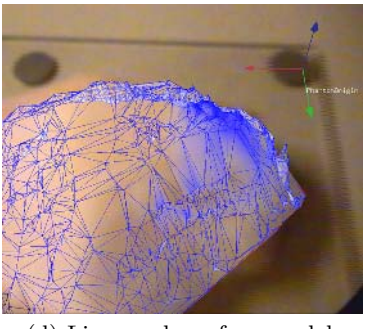


(a) Hardware setup.

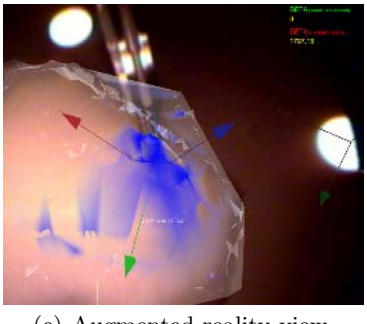
(b) Probe and laparoscope.



(c) Acquired points.



(d) Line mesh surface model.



(e) Augmented reality view.

(f) Virtual probe.

Fig. 3. (a) The hardware setup consists of an optical tracking system (infrared cameras), a tracked camera and a beta-probe. (b) Laparoscopic camera and the extended beta-probe with marker targets for an external optical tracking system. (c)-(e) show screen shots of the system with the acquired data spatially registered into the view of the laparoscopic camera. Optionally all coordinate systems can be visualized during runtime to check the accuracy of the system. (e) visualizes the acquired activity. Blue is high measured activities. The upper number shows the measures of the real beta probe (which was turned off after acquisition), the lower number shows the simulated activity at the tip of the instrument. (f) is the same scene as (e) from outside showing the simulated activity with the 'NodeSeeker' turned off.

for resection. Since we can not assume to have a measure at exactly the same location our instrument is positioned, we use radial basis weights to calculate the predicted activity[5]. The influence I_i of a given point i on the point to be interpolated is assumed to decay exponentially with the squared distance to the point d_i. We used $I_i = \exp\{-d_i^2/(2\sigma^2)\}$ where σ gives a radius of influence of each point in the interpolation. Additionally we defined a cut off radius. Finally the intrapolated/extrapolated measurement $\hat{M}$ was given as the weighted sum of the acquired measurements M_i, that is $\hat{M} = (\sum\ I_i M_i)/(\sum\ I_i)$.

Using the estimated measurement at the tip of the instrument, we either visualize the count rate on a monitor or represent it acoustically in form of a beep (like a traditional Geiger-Mueller counter), just like the one provided by the beta-probe during measurement, which gives extremely intuitive feedback. The codification used is PFM or PTM (pulse frequency or pulse time modulation) with a pulse width of 50 $[ms]$ and an acoustic signal of 800 $[Hz]$. The time between pulses is between 50 and 500 $[ms]$.

4 Experiments and Results

In order to validate the feasibility of the idea and estimate the accuracy of the system, a series of preliminary experiments were designed and performed. For this an *ad-hoc* three dimensional phantom is built. The selected prototype consists of a convex foam volume (e.g. simulating an organ) with attached infrared markers to allow acquisition of data relative to the coordinate system of the phantom and thus rigid motion of the phantom during acquisition and information representation. For the experiments, known volumes of an FDG solution of 2 $[MBq/ml]$ were injected to known positions of the phantom.

The experimental protocol consists in the acquisition of 1000 to 3000 points on a $10 \times 6 \times 4$ $[cm^3]$ volume. The scan time for our phantom was between 2 and 5 $[min]$. However, the scanned degree of detail of the surface was higher than required for surgery, thus the acquisition time can be reduced to about one minute. The scan was performed with the probe approximately orthogonal to the surface of the phantom. In the visual information representation, points are displayed during the acquisition in order to get a visual feed-back of their distribution in space (figure 3(c)). The tip and origin of the probe, the instruments, as well as the origin of the phantom coordinate system can be optionally visualized in the camera view to validate the accuracy of the tracking system and the calibration of probe, instrument and camera during runtime. With the information acquired, the line and triangle meshes generated with the Delaunay algorithm are rendered and visualized superimposed onto the camera view (figure 3(d)). In case of the triangles, the color of the surface and its transparency are encoded as function of the activity measured.

The screen shots shown in figures 3(d) and 3(e), as well as the multimedia additional material[6] show the qualitative performance of the system. The visual registration of three dimensional surface and real surface was achieved up to

[6] Visit http://campar.in.tum.de/BetaProbe for a demo video.

a satisfactory level ($1 - 2$ $[mm]$). Moreover, the navigation system showed hot spots at the locations the radioactivity was injected and showed sufficient spatial accuracy for resection of residual tumor. In the simulated probe modality (figure 3(f)), both the displayed count rate, as well as the acoustic signal, managed to guide the surgical instrument back to spots of high activity.

5 Conclusions

Beta-probes have been extensively studied and applied with success in medicine in the last decade. This is also the case for instrument tracking and advanced visualization in image guided surgery. In the present paper we introduce the integration of both technologies in a navigated beta-probe system for interventional use that benefits from both and allows improved minimal cancer resection. Furthermore, the concepts introduced represent one of the first approaches for three dimensional intraoperative molecular imaging.

The idea of tracking PET probes was mentioned in the past by several groups [10,11,12], but these patents and papers did not provide any theory or application. The combination with augmented reality and virtual measurement is new.

The experiments showed that despite the inherent sources of error in the system like statistical deviation, scattering, tracking inaccuracy, calibration of targets, tips and cameras, etc., the designed system is capable to perform well on the phantom model. In particular, the visual representation was appropriate for the targeted use showing the hot spots at their actual positions and with expected count rates. The simulated probe representation also guided the surgical instrument back to the regions with the highest measured activity.

As part of the future work, animal studies are scheduled. Eventhough no major considerations on deformability, motion and in general artifacts that arise from the clinical application were done, this work presents a new solution which enables the surgeons to perform a resection procedure, which is currently done with minimal support, by providing three dimensional intraoperative molecular imaging and thus by integrating functional information and surgical navigation.

Acknowledgements. The authors want to thank Dr. Markus Schwaiger and Dr. Farhad Daghighian for their support, as well as, Ben Mathison for his detailed technical support.

References

1. Phelps, M.E.: PET: The merging of biology and imaging into molecular imaging. The Journal of Nuclear Medicine **41**(4) (2000) 661–681
2. Levin, C.S., Restom, T.S., Kapelanski, D.P., Hoh, C.K.: A simple surgical probe that directly detects positrons emitted from radiolabeled malignant tissue. Proceedings of the SNM 50th Annual Meeting **44**(5) (2003)

3. Daghighian, F., Mazziotta, J.C., Hoffman, E.J., Shenderov, P., Eshaghian, B., Siegel, S., Phelps, M.E.: Intraoperative beta probe: A device for detecting tissue labeled with positron or electron emitting isotopes during surgery. Med Phys **21**(1) (1994) 153–157
4. Raylman, R.R., Wahl, R.L.: Evaluation of ion-implanted-silicon detectors for use in intraoperative positron-sensitive probes. Med Phys **23**(11) (1996) 1889–1895
5. Weisenberg, A.G.: Gamma-ray blind beta probe. US Patent 6,317,622 (2001)
6. Sielhorst, T., Feuerstein, M., Traub, J., Kutter, O., Navab, N.: Campar: A software framework guaranteeing quality for medical augmented reality. In: Proceedings of CARS 2006 - *to appear.* (2006)
7. Hoff, W.A., Vincent, T.L.: Analysis of head pose accuracy in augmented reality. IEEE Trans. Visualization and Computer Graphics **6** (2000)
8. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. 2nd edn. Cambridge University Press (2003)
9. Daniilidis, K.: Hand-eye calibration using dual quaternions. Journal of Robotics Research **18** (1999) 286298
10. Scibilia, G., Soluri, A.: Surgical probe for laparoscopy or intracavitary tumor localization. US patent 6,021,341 (2000)
11. Daghighian, F., Walker, R.E., Leon, B.J.: Methods and devices to expand applications of intraoperative radiation probes. US patent 6,510,336 B1 (2003)
12. Raylman, R.R., Hyder, A.: A dual surface barrier detector unit for beta-sensitive endoscopic probes. IEEE Trans. Nucl. Sci. **51**(1) (2004) 117–122

Development of Safe Mechanism for Surgical Robots Using Equilibrium Point Control Method

Shinsuk Park, Hokjin Lim, Byeong-sang Kim, and Jae-bok Song

Dept. of Mechanical Engineering, Korea University,
Anam-dong, Sungbuk-gu, Seoul, Korea
{drsspark, jini0117, lovidia, jbsong}@korea.ac.kr

Abstract. This paper introduces a novel mechanism for surgical robotic systems to generate human arm-like compliant motion. The mechanism is based on the idea of the equilibrium point control hypothesis which claims that multi-joint limb movements are achieved by shifting the limbs' equilibrium positions defined by neuromuscular activity. The equilibrium point control can be implemented on a robot manipulator by installing two actuators at each joint of the manipulator, one to control the joint position, and the other to control the joint stiffness. This double-actuator mechanism allows us to arbitrarily manipulate the stiffness (or impedance) of a robotic manipulator as well as its position. Also, the force at the end-effector can be estimated based on joint stiffness and joint angle changes without using force transducers. A two-link manipulator and a three-link manipulator with the double-actuator units have been developed, and experiments and simulation results show the potential of the proposed approach. By creating the human arm-like behavior, this mechanism can improve the performance of robot manipulators to execute stable and safe movement in surgical environments by using a simple control scheme.

1 Introduction

Over the past years, robotic technology has been rapidly gaining acceptance in the operation room. Robots are known to enhance surgery by improving precision, repeatability, stability, and dexterity [1]. Surgical robots are now regularly used in minimally invasive surgery (MIS). Robots for MIS procedures are required to safely perform contact tasks with the patient body in a small workspace.

A robot that contacts the surrounding environment and works within kinematic constraints requires force control, and must be capable of accurately modulating and controlling its actuator torques and forces in addition to knowing where it is in its workspace. Such robots, however, have mainly been limited to laboratory research so far.

Unlike robot manipulators, humans are capable of autonomously performing a wide range of tasks in unpredictable environments. They particularly excel at contact tasks, such as opening a door, turning a knob or lever, and turning a key in a lock. They use a combination of intrinsic compliance, forces and torque sensation, and tactile contact clues from the environments. The human arm muscles are compliant

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 570 – 577, 2006.

actuators that are arranged antagonistically about joints to independently control the multi-dimensional *position* and *compliance* at the limb endpoint. The main characteristic of the human arm is that it introduces relatively high, controllable compliance at the mechanical level: rather than achieving compliance by methods based on controller synthesis, the human arm has *redundant* actuators that provide intrinsic compliance and thus safety [2]. Several studies employed redundant actuators to modulate stiffness of robot manipulators [3],[4].

The purpose of this paper is to present a novel mechanism for surgical robotic systems by adopting the idea of the *equilibrium point control hypothesis*, one of many theories in human motion control. This study proposes to convert the human motor control scheme into a novel robot control scheme using the double-actuator mechanism, which employs two actuators for each joint: one for position control and the other for stiffness control. A two-link manipulator and a three-link manipulator with the double-actuator units have been developed, and experiments and simulation were conducted to show the validity of this mechanism in surgical environments.

2 Double-Actuator Mechanism

While a great number of studies have attempted to reveal the basic properties of the human neuromuscular control system, understanding the actual control strategies adopted by the CNS is a fundamental problem of neurophysiology yet to be explored [5],[6]. A simple approach to this problem suggests that the CNS utilizes the effective dynamic and mechanical properties of the muscles and neural feedback circuits to circumvent the computational complexities. For a single joint, the combined action of a group of muscles spanning the joint, both agonists and antagonists, determines an equilibrium posture for the joint. Central command generates a series of equilibrium points for a limb, and the "spring-like" properties of the neuromuscular system will drive the motion along a trajectory that follows these intermediate equilibrium postures. This *equilibrium point control hypothesis* applies to the control of both static posture and dynamic movement [7]. In the framework of equilibrium point control, the resultant joint torques are determined only by deviation of the actual trajectory from the equilibrium point trajectory and joint impedance.

It should be noted that equilibrium point control applies the same strategy to tasks requiring interaction with the environment, unrestrained motions and the transition between the two. Control of contact force can be achieved through the use of an equilibrium point, as well as the controlled joint stiffness: simply moving the equilibrium point to a point within a contact object will cause the limb to exert a force on that object.

One way to implement the control scheme would be placing the *positioning actuator* and the *stiffness modulator* in series on one joint axis. In this study we employed an electrical motor with a higher gear ratio as the positioning actuator, and an electrical motor with a lower gear ratio as the stiffness modulator. Fig. 1 (a) illustrates the two actuators placed on one joint axis. With the mechanism, we can control the equilibrium point and joint stiffness at the same time: one actuator commands the equilibrium point (positioning actuator), the other joint stiffness (stiffness modulator). The positioning actuator provides the equilibrium point (angle), and the stiffness modulator controls the

restoring torque around the point. The positioning actuator is fixed to the base link, and the moving link is fixed to the output axis of the stiffness modulator. Fig. 1 (b) shows an equivalent mechanism to the one in Fig. 1 (a). Two actuators are connected in parallel by using a gear train: one gear is fixed to the output axis of the positioning actuator, the other to the frame of the stiffness modulator.

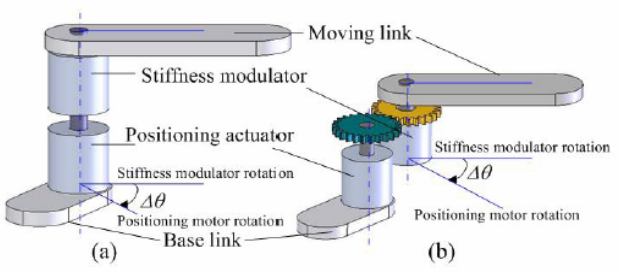


Fig. 1. Conceptual figures of two equivalent double actuator mechanisms (a) two actuators connected in series, and (b) two actuators connected in parallel using a gear train

Fig. 2 illustrates the double-actuator mechanism we have developed based on the idea of Fig. 1 (b). For the positioning actuator and stiffness modulator of the manipulator, BLDC motors were used, where the stiffness modulator had a lower gear ratio for back-drivability when attempting compliant motion control using the proposed control scheme.

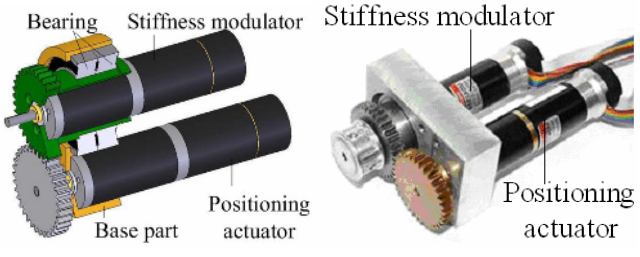


Fig. 2. Double-actuator mechanism with two parallel actuators

3 Manipulators with Double-Actuator Units

When the manipulator equipped with the double-actuator mechanism is in contact with an object, the contact force can be controlled either by moving the equilibrium point further into the contacting object or by modulating the stiffness at the endpoint of the manipulator. For a desired endpoint stiffness matrix $\mathbf{K}_x$ in the Cartesian coordinate system, the joint stiffness matrix $\mathbf{K}_J$ of the stiffness modulator can be calculated using the manipulator Jacobian $\mathbf{J}_M$ as follows:

$$\mathbf{K}_J = \mathbf{J}_M{}^T \mathbf{K}_x \mathbf{J}_M \tag{1}$$

The contact force **F** at the endpoint can be simply estimated by using the following relationship:

$$\mathbf{F} = \mathbf{J}_M{}^{-T} \boldsymbol{\tau} = \mathbf{J}_M{}^{-T} (\mathbf{K}_J \Delta\boldsymbol{\theta}) = \mathbf{J}_M{}^{-T} (\mathbf{J}_M{}^T \mathbf{K}_x \mathbf{J}_M) \Delta\boldsymbol{\theta} = \mathbf{K}_x \mathbf{J}_M \Delta\boldsymbol{\theta} \tag{2}$$

where $\boldsymbol{\tau}$ is the torque exerted by the stiffness modulator and $\Delta\boldsymbol{\theta}$ is the difference in joint angles between the equilibrium posture and actual posture. Since the stiffness modulator is capable of *active* stiffness modulation based on angle measurements, the joint stiffness matrix $\mathbf{K}_J$ realized by the stiffness modulator contains off-diagonal terms, unlike *passive* compliant elements discussed in introduction.

3.1 Manipulation of Stiffness Ellipse Using Two-Link Manipulator

Fig. 3 shows a two-link manipulator equipped with the double actuator mechanism described in the previous section. The double-actuator units that drive the first and the second link are attached to the base and the first link, respectively. The driving torque of the double-actuator units were transferred to the links by timing pulleys and timing belts. Encoders were equipped to the output axis of each actuator to obtain position feedback. With the two-link manipulator, stiffness matrices at the endpoint can be manipulated arbitrarily in 2D space. Simply by manipulating stiffness ellipse, we can easily control the magnitude and direction of the contact force between the robot and the environment. Also, the double-actuator mechanism allows us to estimate contact force only by the angle measurements at the joints, without using a force transducer at the endpoint.

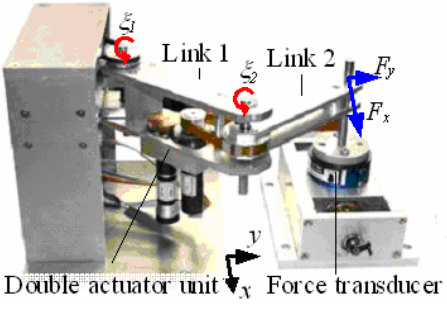


Fig. 3. Two-link manipulator with double actuator mechanism

Experiments. With a given endpoint stiffness matrix in Cartesian coordinate system, force-displacement relation at the endpoint was measured by using a force transducer (JR3 J409). With the given stiffness matrices, the equilibrium point of the endpoint was moved in six different directions so that force-displacement relation can be measured in the directions. In this experiment, three stiffness matrices were given for three different configurations of the manipulator. Fig. 4 compares the target stiffness

ellipses and the estimated stiffness ellipses. As can be seen in the figure, the actual stiffness is nearly identical to the target stiffness in all the configurations.

Fig. 5 compares the measured and estimated contact forces while the endpoint is in contact with an object. The estimated force was based on encoder data from the two double-actuator units in the two-link manipulator. The figure demonstrates that the estimated force is very close to the measured force.

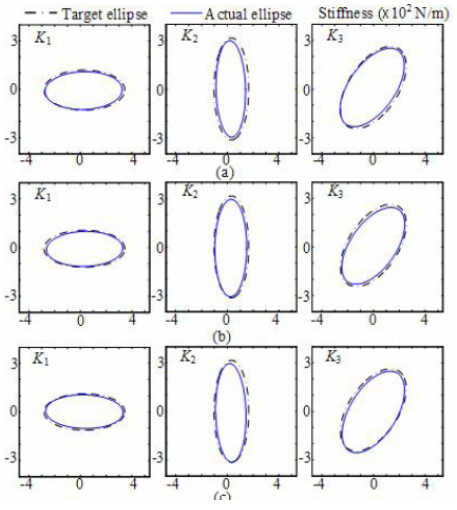


Fig. 4. Comparison of target and actual stiffness ellipse

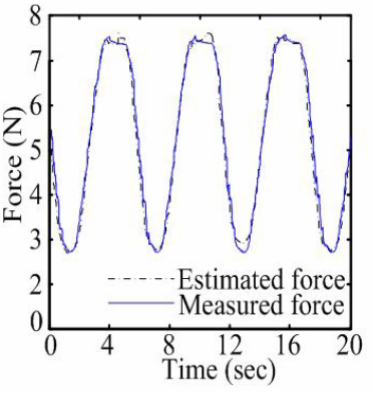


Fig. 5. Comparison of measured and estimated force

3.2 Adaptive Control of Remote Compliance Center Using Three-Link Manipulator

In MIS, surgical procedures are carried out using long instruments inserted into the patient's body through small incisions with the surgeon manipulating the tool handles from outside the patient's body. The remote center of motion (RCM) that constrains the motion of the instrument through a known fulcrum point has been used in several laparoscopic robots, including commercial systems such as the da Vinci (Intuitive surgical) [8]. With the RCM method, however, a small error in the fulcrum position can cause a large contact force between the patient body and the instrument.

Remote compliance center (RCC) devices have long been used to facilitate insertion for assembly operations in manufacturing [9]. By decoupling lateral and angular motion around the compliance center, a RCC device can ease the contact force during insertion of a part into a hole. When a project compliance center is near the insertion point, the part translates into the hole when it strikes the outside without rotating. This translation without rotation prevents the jamming and galling, which can generate excessive contact force. Since conventional RCC devices employ passive mechanical compliance, the compliance center is fixed at the tip of the part to be inserted and moves further into the hole after the part is inserted.

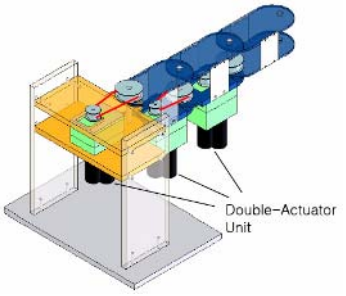


Fig. 6. Three-link manipulator with double actuator mechanism

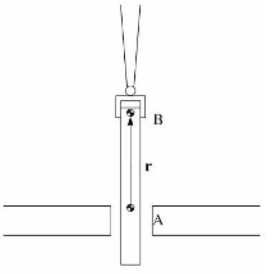


Fig. 7. Insertion of an instrument into a port

By implementing three double-actuator units on a three-link manipulator, the position of the compliance center can be controlled in two-dimensional space so that the compliance center is maintained near the insertion point during the whole insertion operation (Fig. 6). This adaptive control of RCC, when applied to MIS, can be advantageous, since it prevents excessive force at the port site where the surgical instrument or endoscope is inserted. Fig. 7 illustrates the insertion of a surgical instrument into a port. The position of the endpoint of the robot manipulator (point B) with respect to the insertion point (point A) is represented by vector $\mathbf{r}$ ($\mathbf{r} = (r_x, r_y)^T$).

Using the double-actuator mechanism, the compliance center can be placed at the insertion point (point A), and the compliance matrix described at point A is given as:

$$\mathbf{C}_A = \begin{bmatrix} C_{xx} & C_{xy} & 0 \\ C_{yx} & C_{yy} & 0 \\ 0 & 0 & C_{\theta\theta} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_r & \mathbf{0} \\ \mathbf{0} & C_\theta \end{bmatrix} \tag{3}$$

The compliance matrix is transformed by Jacobian matrix $\mathbf{J}$ between point A and point B to yield the compliance matrix at the endpoint of the manipulator (point B).

$$\mathbf{C}_B = \mathbf{J}\mathbf{C}_A\mathbf{J}^T = \begin{bmatrix} \mathbf{C}_r + \mathbf{r}^*\mathbf{C}_\theta\mathbf{r}^{*T} & \mathbf{r}^*\mathbf{C}_\theta \\ \mathbf{C}_\theta\mathbf{r}^{*T} & \mathbf{C}_\theta \end{bmatrix} \tag{4}$$

where $\mathbf{r}^* = (-r_y, r_x)^T$.

With the predetermined compliance matrix $\mathbf{C}_A$ at the insertion point and the vector $\mathbf{r}$ that can be estimated from forward kinematics of the manipulator, we can calculate the compliance matrix $\mathbf{C}_B$ .The joint stiffness $\mathbf{K}_\theta$ for the manipulator can be determined using the manipulator Jacobian $\mathbf{J}_M$:

$$\mathbf{K}_\theta = \mathbf{J}_M{}^T\mathbf{C}_B{}^{-1}\mathbf{J}_M \tag{5}$$

Simulations. Using a commercial dynamic simulation software (MSC Visual Nastran), we performed simulations to compare the performance of the adaptive RCC

method with conventional RCC devices. Fig. 8 compares the contact impulse during insertion of a long instrument into a port. While the first peak vales are nearly the same, the second peak impulse with the adaptive RCC control is lower than that of conventional RCC. Fig. 9 compares the contact impulse while the instrument is moving laterally while inserted halfway into the hole. Compared to conventional RCC, we can significantly reduce the contact force with the adaptive RCC control.

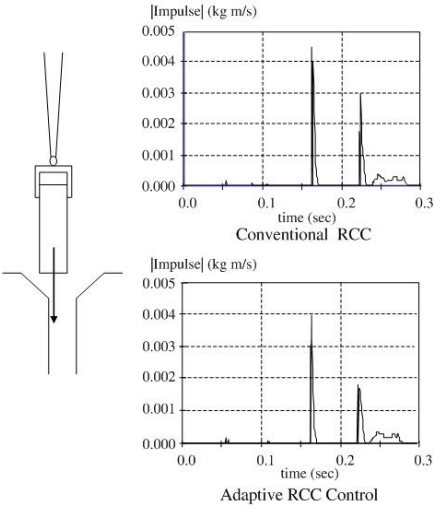


Fig. 8. Comparison of contact impulse during insertion

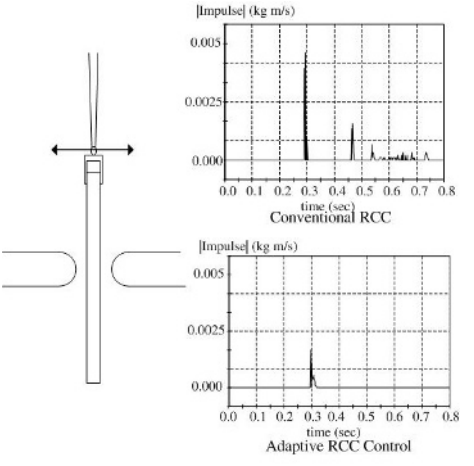


Fig. 9. Comparison of contact impulse during lateral motion

4 Conclusions

This work proposes a novel mechanism for robot manipulators based on the equilibrium point control scheme. Experimental results demonstrated that the mechanism could successfully control position and compliance of the robot under the equilibrium point control scheme.

There are several benefits of the proposed control scheme. One of the benefits would be the simplicity of control. Equilibrium point control applies the identical strategy from motion in free space to contact task. In cases when in contact, the contact force can be applied by simply moving the equilibrium point within the contact object. Contact force can also be tailored by tuning the compliance of the second motor. Secondly, equilibrium point control turns the force control problem into a position control problem. In this control scheme, output torque is proportional to the difference in the angles, where the angles are much easier to sense and control than the torques. In many force control schemes, force/torque sensors are equipped near the end-effector to detect the contact force at the end-effector. It is difficult, however, to detect contact on the parts far from the end-effector. The developed mechanism estimates the contact force based on the angle measurement at the joints. Thus, the contact on the part far from the end-effector can also be detected and estimated.

A conceivable application of the developed mechanism would be teleoperation in surgical environments, where robotic systems are required to regenerate the movement of the surgeon while regulating the contact force under a safe level. There are several issues to be considered, including the size of mechanism and the extension of DOFs. These issues will be studied in our future study.

Acknowledgment

This work was supported by grant No. (R01-2003-000-10336-0) from the Basic Research Program of the Korea Science & Engineering Foundation.

References

[1] Howe, R., Matsuoka, Y.: Robotics for Surgery. Ann. Rev. Biomed. Eng. (2000) 1: 211-240.
[2] Tonietti, G., Schiavi, R., Bicchi, A.: Design and Control of a Variable Stiffness Actuator for Safe and Fast Physical Human/Robot Interaction. Proc. of IEEE Int'l Conf. on Robotics and Automation (2005) 528-533.
[3] Yi, B., Freeman, R., Tesar, D.: Open-loop Stiffness Control of Overconstrained Mechanisms/Robotic Linkage Systems. Proc. of IEEE Int'l Conf. on Robotics and Automation (1989) 1340-1345.
[4] Yi, B., Freeman, R., Tesar, D.: Geometric Characteristics of Antagonistic Stiffness in Redundantly Actuated Mechanisms. Proc. of IEEE Int'l Conf. on Robotics and Automation (1993) 654-661.
[5] Bernstein, N.: The Coordination and Regulation of Movements. Pergamon Press (1967).
[6] Arbib, M.: The Metaphorical Brain: An Introduction to Artificial Intelligence and Brain Theory. Interscience (1972).
[7] Bizzi, E., Hogan, N., Mussa-Ivaldi, F., Giszter, S.: Does the Nervous System Use Equilibrium-Point Control to Guide Single and Multiple Joint Movements. Behavioral and Brain Sciences, Vol. 15 (1992) 603-613.
[8] Taylor, R., Funda, J., Eldridge, B., Gruben, K., LaRose, D., Gomory, S. Talamini, M., Kavoussi, L., Anderson, J.: A Telerobotic Assistant for Laparoscopic Surgery. IEEE EMBS Magazine Special Issue on Robotics in Surgery (1995) 279-291.
[9] Whitney, D.: Part Mating in Assembly. Handbook of Industrial Robotics (1985) 1084-1116.

Real Time Adaptive Filtering for Digital X-Ray Applications

Olivier Bockenbach[1], Michel Mangin[2], and Sebastian Schuberth[1]

[1] Mercury Computer Systems, GmbH, Lepiusstr. 70,
12163 Berlin, Germany,
[2] Mercury Computer Systems, Sarl, 19 Avenue du Quebec,
91951 Courtaboeuf, France
{obockenbach, mmangin, sschuber}@mc.com

Abstract. Over the last decade, many methods for adaptively filtering a data stream have been proposed. Those methods have applications in two dimensional imaging as well as in three dimensional image reconstruction. Although the primary objective of this filtering technique is to reduce the noise while avoiding to blur the edges, diagnostic, automated segmentation and surgery show a growing interest in enhancing the features contained in the image flow.

Most of the methods proposed so far emerged from thorough studies of the physics of the considered modality and therefore show only a marginal capability to be extended across modalities. Moreover, adaptive filtering belongs to the family of processing intensive algorithms. Existing technology has often driven to simplifications and modality specific optimization to sustain the expected performances.

In the specific case of real time digital X-ray as used surgery, the system has to sustain a throughput of 30 frames per second.

In this study, we take a generalized approach for adaptive filtering based on multiple oriented filters. Mapping the filtering part to the embedded real time image processing while a user/application defined adaptive recombination of the filter outputs allow to change the smoothing and edge enhancement properties of the filter without changing the oriented filter parameters.

We have implemented the filtering on a Cell Broadband Engine processor and the adaptive recombination on an off-the-shelf PC, connected via Gigabit Ethernet. This implementation is capable of filtering images of 5122 pixels at a throughput in excess of 40 frames per second while allowing to change the parameters in real time.

Keyword: Adaptive Filtering, Digital X-Ray, Cell Processor, Real Time, Medical Imaging.

1 Introduction

1.1 Background

Quality becomes one of the most important metrics when a radiologist or a surgeon examines the results from medical imaging. Medical imaging needs to provide the observer with images where the features of interest are highlighted while the interfering

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 578 – 587, 2006.

noise is reduced. Over the last decades, the medical imaging community has passionately debated on the approaches and results for noise reduction and edge enhancement. Because image quality has been expressed as resolution and Signal to Noise Ratio (SNR), analyzing the physics of the imaging system and the properties of received signal from which image data is sampled has lead to implementation of spatial and temporal filters. The goal of these two categories is to take advantage of two properties of noise. Its temporal randomness is such that it zeroes itself out over time and its low frequency spectrum allows to use simple low pass filters to smooth it locally out.

Because the constant effort it to reduce the radiation dose during examination, the complexity in the implementation of this type of filter has been growing accordingly, moving from methods where the filter would be uniformly applied to a complete image set to more data dependent filters. Recently, adaptive filtering methods have been introduced[3,4].

Adaptive filtering can be defined as a method that takes advantage of a priori knowledge of the properties of the signal in the neighborhood of the location to be filtered. Most sophisticated methods proposed so far take advantage of the high frequency spectrum to keep the information about the underlying structures (such as edges) and use the low frequency components to identify the parameters of the noise. Most recent works use a combination of High Pass (HP) and Low Pass (LP) filters to determine the way the input data should be filtered. Such methods are extremely compute intensive since the computation requires to evaluate the output of several filters before any filtering can be applied. Even on common high-end hardware, the cost issues restrict designers to compromise and to optimize the filter for a given problem, indeed limiting the generality of the solution. Therefore, although those methods show extremely good results, they suffer from various shortcomings that prevent a wide adoption in other modalities than those for which they have been designed for[6,7].

Real-time digital X-ray adds a stability issue to the filtering problem. In fact, when the observer looks at a sequence of images and concentrates on certain features in the display, those features need to be filtered and displayed in a consistent way, e.g. variations in the grey level should not appear for obvious reasons. This is why the adaptiveness of the filter needs to be constrained to some basic parameters to avoid negative effects for the observer, such as flickering. Furthermore, post-processing algorithms such as automated segmentation and registration work best when previous processing steps maintain temporal consistency in pixel values.

1.2 Goal of the Study

In this study, we start from a generic adaptive algorithm that can theoretically address all types of de-noising and edge enhancements endeavors. It notably consists of a set of HP and LP filters as well as a recombination function. We show that is possible to:

- identify a set of specific parameters that apply only to the filter components and that allow to cover specific Digital X-ray examinations,
- precompute the filter kernels based on these parameters so that filtering stability is insured for as long as needed,
- adapt the recombination of the filters output in such a way that an application or an operator can easily select the contribution of the different filters from simple ergonomic settings.

The idea behind this algorithm is to separate the noise from the rest of the image by trying to find a structure in the sampled data by assuming that random noise does not follow that structure and that anything without structure can be smoothly extracted from the input data. As appealing as this scheme is for areas with low frequency variations, this dichotomy does not work well in the general case where noise appears in conjunction with structured features such as tissues and bones. The separation of noise from tissue requires more sophisticated methods. We have chosen to locally track the structure of the underlying features and more specifically, their orientation. Similar work was proposed by Raupach et all[1] in the 3D case.

To give an overview, section 2 of this document describes the structure of the filter and the way the parameters influence the processing. A detailed mapping on the hardware is also described in this section. Section 3gives a short description of the data set. For this document, we have selected static images from different X-Ray examination. All images are taken with low radiation dose. In section 4, filtered images are shown and we discuss the results with respect to the parameter used.

2 Methodology

2.1 Principles

In this section we show how to evaluate the structure of the features in the images and steer the way the filter operates. Two main properties are combined in the filtering process : the image gradient and the local orientation of structures. In areas with low gradients, the filter operates like a smoothing function reducing noise while in areas with larger gradients, the filter becomes more an identity function, preserving the high frequency information given by the edges. In the areas with larger gradients, it is important to evaluate the dominant direction in which the filter shall operate. This main direction defines the level of local anisotropy of the filter. In areas where there is no major direction, the filter turns into an isotropic filter.

In other areas, there may be fast and strong orientation changes in the main direction due to fading in and of main structures. The filter direction must smooth out the local inconsistencies of the filter direction. A local Gaussian filter can be used for that purpose.

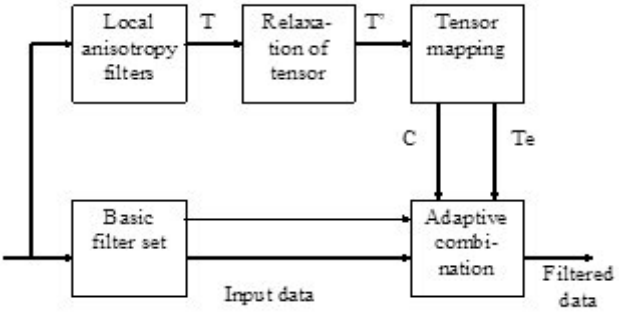


Fig. 1. Structure of the adaptive Filter

2.2 Filter Implementation

The logical design of our implementation is based on the architecture of Westin et all[4] depicted in figure 1. Although the structure seems very similar, the filter presented in this document defers in the type of the filters and their parameters.

The images to be filtered follow two different paths. For the Basic Filter Set, we have opted for a simple filter LP given by equation [1]. The complementary Band Pass (BP) filters are based on equation [2] and oriented along three directions, regularly distributed on the interval [0, π]. first path gives the estimation of the local anisotropy.

$$Flp(W) = \begin{cases} \cos^2(\frac{\pi W}{2Wc}) & \text{for } W \le Wc \\ 0 & W > Wc \end{cases} \quad (1)$$

$$Fhp(W) = \begin{cases} 1 - \cos^2(\frac{\pi W}{2Wc}) & W \le Wc \\ 1 & \text{for } Wc < W \le \pi - Wc \\ \cos^2(\frac{\pi(W - \pi + Wc)}{2Wc}) & \pi - Wc < W \le \pi \end{cases} \quad (2)$$

The anisotropic filters are derived from band pass lognormal filters as given in equation [3], with a 6dB of sensitivity bandwidth B and center frequency W_0. They are oriented in the same directions at those used in the basic filter set.

$$H_\rho(W) = e^{-4\ln^2(\frac{W}{W_0})/(B^2\ln(2))} \quad (3)$$

The processing of the anisotropic filters is followed by a relaxation step and a tensor mapping step, this path provides the information that will control the use the basic filter set..

We have implemented all filters in the spatial domain with a convolution function. The filter kernel size has been set to 15*15. It is initialized with the values depending on the type of the filter, the central frequency W_0 and the cut off frequency W_c. The output of the filters is represented with tensors.

The tensor relaxation step is implemented as a simple convolution with Gaussian filter. Its kernel size is again 15 * 15.

In the general case, the tensor mapping steps requires the evaluation of the Eigenvalues of the tensors. However, since we are in the two dimensional case, we have used the direct mathematical solution for the Eigenvalues instead of using more complex iterative methods. The output of this step at location x is driven by parameters (α, a, b, c, λ_1) defining the trade off between resolution and stability, as given in equation 4.

$$o(x) = c(1 - \lambda_1') + \frac{\lambda_1'^{b}}{\lambda_1'^{a+b} + \alpha^b} \quad (4)$$

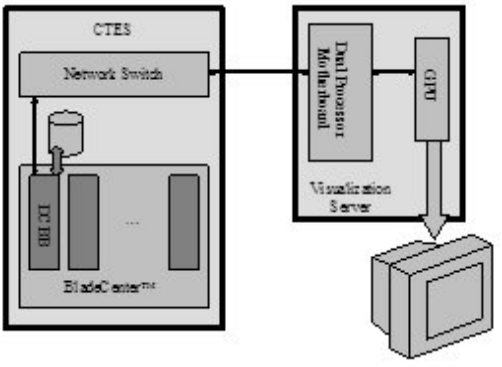


Fig. 2. System Block Diagram

Where :

$$\lambda_1' = \frac{\lambda_1}{\lambda_{1\max}} \tag{5}$$

and where $\boldsymbol{\lambda_{max}}$ is the largest Eigenvalue of the tensor. $\boldsymbol{\lambda_1}$ is the largest eigenvalues of the tensor while the parameters **a**, **b** and **c** are user definable and arguments to the tensor mapping step. It is worth noticing that an output value larger than 1 increases the high-frequency components of the input image and hence acts as an edge enhancement feature in the filtering process.

The final adaptive filter step then merely consists of a linear combination of the output of the basic filters weighted by the control tensor resulting from the tensor mapping step.

2.3 Hardware

The system hardware prototype is based on a Mercury Cell Technology Evaluation System (CTES) which consists of a BladeCenter™ [1] hosting one Dual Cell-Based Blade (DCBB)[2] For convenience, we connected the DCBB to a visualization server via Gigabit Ethernet (GE), intended to run the control part of the visualization application.. The relationship between these components is shown in Fig. 2.

The visualization can be considered as a standard PC and runs an off-the-shelf Windows XP[3] operating system. It provides all the required means for network communications and supports the drivers for the selected GPU board. In addition, it also enables us to run Amira[4] 3.1.1 for real-time visualization of the reconstructed slices and for volume rendering.

The DCBB is a board built around a dual Cell Broadband Engine (CBE) architecture and designed to be collocated with other boards, potentially other DCBBs, in a BladeCenter. DCBB boards feature two CBE processors in Symmetric Multi Processor

[1] International Business Machines Corporation, Armonk, New York, USA.
[2] Mercury Computer Systems, Inc., Chelmsford, Massachusetts, USA.
[3] Microsoft Corporation, Redmond, Washington, USA.
[4] Mercury Computer Systems, Inc., Chelmsford, Massachusetts, USA.

(SMP) mode, clocked at 3 GHz (Fig.3). Each CBE processor has 512 MB of Rambus (XDR) memory directly connected to its pins A sound description of the CBE processor has been given by Pham et al[8] .

2.4 Algorithm Mapping

We have implemented the major part of the processing on one of the CBE processors of the DCBB. It receives the image flow over Gigabit Ethernet at a throughput rate of 30 frames per second. The images are encoded as 16 bits integers. They are first converted to floating point values on the PPE before being distributed over the SPEs.

The processing step on the DCBB are implemented in such a way that they produce for output the norm of the filtered images, the LP component and the weighted sum of the HP filters along with important resulting values such as the largest Eigenvalues. The output images are sent over the network to the visualization server where the smoothing and the adaptive combination take place.

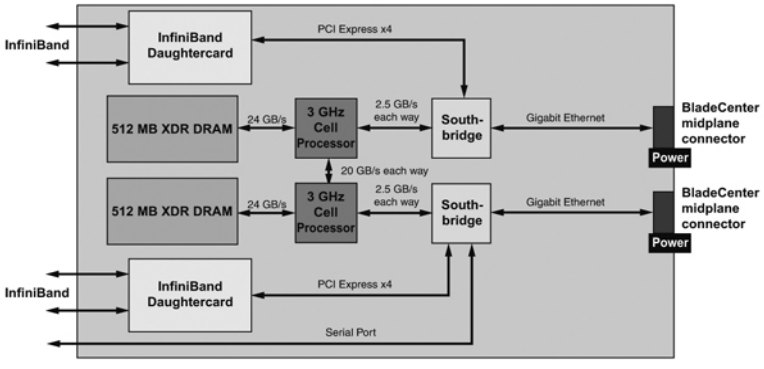
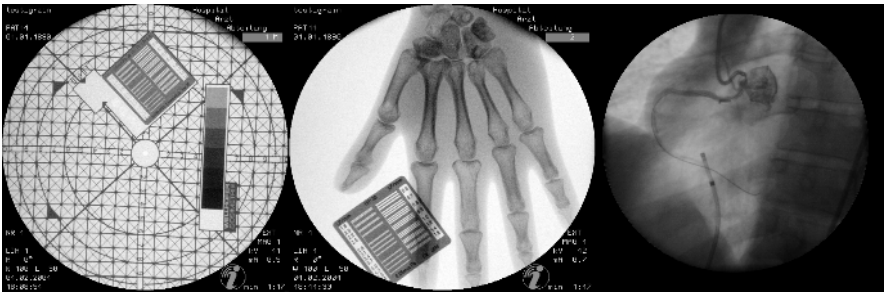


Fig. 3. Dual Cell Based Board

The application running on the visualization server captures the user settable values for α, **a**, **b** and **c** in equation [4], perform the computation of the resulting images accordingly. Finally the images are sent to the monitor and other post-processing steps.

3 Data Sets

To evaluate our implementation, we have selected two different data sets. The first one consists of two low-dose images (Fig 4a and 4b), along with a dosimeter and a resolution grid. We used those images to set the central and cutting frequencies of the filters and evaluate their effects on the noise and on the enhancement of features. With those parameters identified we have run the filter process against several low-dose cardiac sequences from which a sample image is shown is Fig 4c.



a: Low-dose dosimeter – **b:** Hand with resolution grid – **c:** cardiac sequence

Fig. 4. Samples from the data sets

4 Results and Discussion

Fig. 5 shows a typical result obtained when using this adaptive filter with three filter directions. Tests with a higher number of directions have been made without showing significant improvement in the image quality; the differences between images filtered with three and with seven direction show only minor effects in the presence of edges.

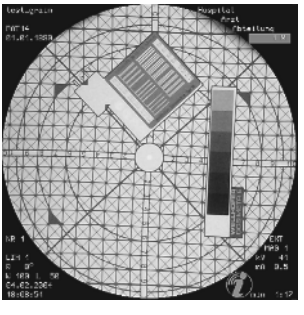
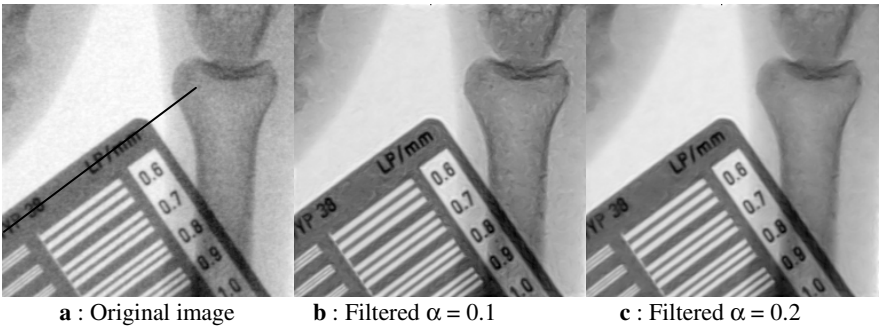


Fig. 5. Image 4-a filtered

The tensor mapping uses the following parameters $\alpha = 0.1$, $\mathbf{a} = 0.5$, $\mathbf{b} = 2$ and $\mathbf{c} = 0.25$.

The image shows that the noise on the dosimeter has significantly gained in homogeneity and the contrast has improved the definition of the lines. There is no blurring at the edges and the overall impression is that the filtered image is of better quality.

Figure 6 represents a part of figure 4-b. 5-a is a part of the original image, unfiltered with a significant level of noise. Nevertheless, the features can clearly be identified by the observer. On the other hand, extracting features by automated segmentation

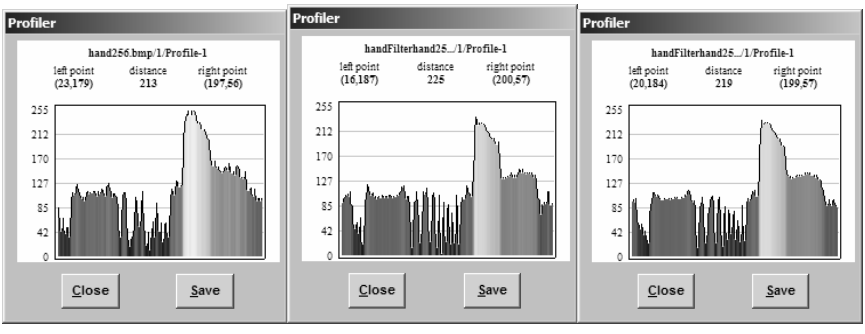


a : Original image **b** : Filtered $\alpha = 0.1$ **c** : Filtered $\alpha = 0.2$

Fig. 6. A part of image 4-b, unfiltered and filtered

may become challenging as one can see in the upper right corner of the resolution grid. Indeed the borders of the letters "LP/mm" are slightly blurred. Automated segmentation may have difficulties with smoother edges such as large bones.

Figures 6b and 6c represent filtered versions of the original image 6-a, with respectively the filter parameters {$\alpha = 0.1$, **a** = 0.5, **b** = 2, **c** = 0.25} and {$\alpha = 0.2$, **a** = 0.5, **b** = 2 and **c** = 0.25}. Both images 6b and 6c show a significant reduction of the noise. Figure 6b also shows a stronger structure of the image. Even though that structure exists in figure 6a, it has been enhanced to a point where it disturbs the image analysis. Automated segmentation is tempted to find edges in areas where the filter has enhanced them. Figure 6c demonstrates the effect of parameters **a**, which drives the contribution of the LP filters. It can be clearly seen that there are significantly less undesired enhancement of structures in figure 6c than in 6b. The first experiments with the interactive segmentation module of Amira demonstrate better identification of the features in the image 6c than on 6b.



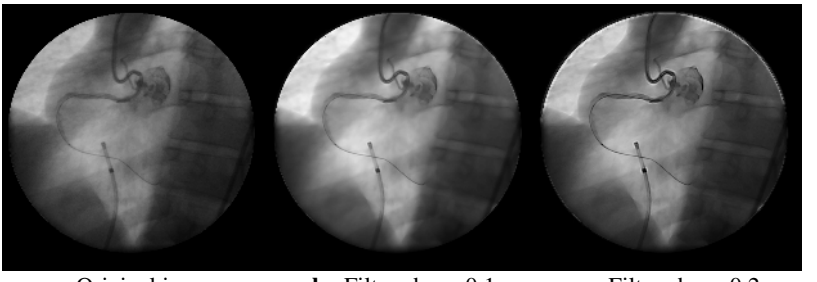
a : Original image **b** : Filtered $\alpha = 0.1$ **c** : Filtered $\alpha = 0.2$

Fig. 6d. Profiles along the black line in figure 6a

Figure 6d shows the data profile along the black line shown in fig. 6a. One can see that the level of noise is lower in the filtered images and that the contrast has been enhanced. It is also worth noticing that the resolution has not been affected.

We have also applied the filters on cardiac images. Figure 7a shows the original image taken from a sequence of 25 images. We have applied the same parameters to the filter as in the previous experiment. We obtained the images shown in 7b and 7c. It is clear to see that 7b smoothes out a significant part of the noise and enhances the dynamics of the image. What cannot be identified on those images but is clearly visible on the real-time display of the sequence is that the structure that the filter has enhanced low-density tissue that could have been taken as noise uniformities. Figure 7c shows a significant enhancement of the contrast in all structures, still without blur. The catheter is clearly visible as well as the boundaries of the bubble.

In the case of the cardiac sequence, we have loaded the 25 images into the Cell's XDR memory and performed the filtering on those images. The current implementation showed that we could sustain 41 slices/sec for images of 1024*1024 pixels. The filtered images were sent to the visualization server where the remaining parts of the adaptive combination of the filter outputs were combined. We could sustain a display rate of 50 frames per second.



a : Original image **b** : Filtered $\alpha = 0.1$ **c** : Filtered $\alpha = 0.2$

Fig. 7. Image 4-c, unfiltered and filtered

The very promising and important result of this study is that it seems possible to create :

- a set of filters that defines the variables to control the adaptive filtering of a sequence of images,
- a set of parameters that make the combination of the filter outputs in such a way the resulting image combines a user settable mixture of noise reduction and edge enhancement.

Furthermore, today's multi-parallel technology can carry out the computation in real-time. There are also direct implications in the proposed architecture of this adaptive filtering. Firstly, the big burden of the processing is performed by one computer which can be embedded into the existing X-Ray cabinet and the visualization of the images only consists of a relatively light processing step. It is hence possible to get several different monitors that display the images with other filtering parameters. This makes it also possible to direct the images flow to other automated post-processing application with specific filtering requirements while keeping a visualization monitor with user friendly settings.

5 Future Work

The work on the implementation of adaptive filtering will follow several directions. The first one is a direct application of the results we have obtained today to automated segmentation and computer vision. We have archived some preliminary results with this adaptive filtering technique that show interesting aspects in this area. Figure 8 displays another experiment where we have put the tensor mapping parameters to $\{\alpha = 0.4$, $\mathbf{a} = 2$, $\mathbf{b} = 2$ and $\mathbf{c} = 4\}$. The picture shows extreme structuring of the features of the image. It is also remarkable that features with very high contrast compared their neighborhood are even more accentuated. This is typically the type of filtering that computer vision applications require. Applying a simple threshold on such an image allows to extract the most important features that significantly simplifies the positioning problem in robotic surgery.

Fig. 8. Image 4-a filtered

The second axis of research consists in applying this technique to the filtering of projection data in the spiral CT applications as well as the filtering of the 3D volumes resulting from the CT reconstruction process.

On the pure technical side, the filtering process will include a multi-resolution pyramid. The expected outcome from this modification is yet another improvement in the image quality. Keeping a constant image quality, we would be capable of reducing the convolution kernel size. This processing burden reduction will have a direct impact on the throughput and make it possible to sustain higher image rates and handle larger images.

Acknowledgements

The authors gratefully acknowledge the help from the Performance Algorithm Group from Mercury Computer Systems for its support in programming the Cell Broadband Engine processor.

References

1. R Raupach, H Bruder, K Stierstorfer, C Suess, T Flohr, *A novel approach for efficient edge preserving noise reduction in CT volume data*, RSNA 2005, Paper SSQ16-03
2. J. Gurung, M. Fawad Khan, A. Maataoui, C. Herzog, R. Bux, H. Bratzke, H. Ackermann, T.J. Vogl, *Multislice CT of the pelvis : dose reduction with regard to image quality using 16-row CT*, Eur. Radiol. (2005) 15:1898-1905
3. H Öktem, K Egiazarian, J Niittylahti, J Lemmetti, *An Approach to Adaptive Enhancement of Diagnostic X-Ray Images*, EURASIP Journal on Applied Signal Processing 2003:5,430-436
4. C F Westin, J Richolt, V. Moharir, R Kikinis, *Affine adaptive filtering of CT data*, Medical Image Analysis 4 (2000) 161-167
5. L Haglund, *Adaptive Multidimensional Filtering*, PhD Dissertation, University Linköping, Sweden
6. K F Westin, L Wigström, T Loock, L Svöqvist, R Kikinis, H Knutsson, Three Dimensional Adaptive Filtering in Magnetic Resonance Angiography, Journal of Magnetic Resonance Imaging 14:63-71 (2001)
7. M Kachelriess, O Watzke, W Kallender, Generalized multi-dimensional adaptive filtering for conventional and spiral single slice, multi slice and cone beam CT, Med. Phys.18 (4), April 2001.
8. D.Pham, S. Asano, M. Bollinger, M.N. Day, H.P. Hofstee, C. Johns, J. Kahle, A. Kameyama, J. Keaty, Y. Masubuchi, M. Riley, D. Shippy, D. Stasiak, M.Suzuoki, J, Warnock, S. Weitzel, D. Wendel, T. Yamazaki, K. Yazawa, *The Design and Implementation of the First-Generation CELL Processor*, ISSCC 2005.

Semiautomatic Volume Conductor Modeling Pipeline for Imaging the Cardiac Electrophysiology Noninvasively

Bernhard Pfeifer[1], Michael Seger[1], Christoph Hintermüller[1], Gerald Fischer[1], Friedrich Hanser[1], Robert Modre[2], Hannes Mühlthaler[3], and Bernhard Tilg[1]

[1] Institute of Biomedical Engineering, University for Health sciences, medical informatics and technology, Austria
bernhard.pfeifer@umit.at
[2] ARC Seibersdorf research GmbH, Austria
[3] Department of Vascular Surgery, Innsbruck Medical University, Innsbruck

Abstract. In this paper we present an approach for extracting patient individual volume conductor models (VCM) using volume data acquired from Magnetic Resonance Imaging (MRI) for computational biology of electrical excitation in the patient's heart. The VCM consists of the compartments chest surface, lung surfaces, the atrial and ventricular myocardium, and the blood masses. For each compartment a segmentation approach with no or little necessity of user interaction was implemented and integrated into a VCM segmentation pipeline to enable the inverse problem of electrocardiography to become clinical applicable. The segmentation pipeline was tested using volume data from ten patients with structurally normal hearts.

1 Introduction

Computer modeling of bioelectric activity is of interest in cardiac electrophysiology [1,2] and, furthermore, the noninvasive imaging of the electrical excitation from electrocardiogram (ECG) becomes more and more a diagnosis tool in clinical application [3,4]. For enabling noninvasive imaging four dimensional (3D and time) anatomical and ECG mapping data have to be combined. The anatomical data are obtained via MRI, and are the basis for generating a VCM. A VCM for solving the inverse problem consists of the electrostatically most relevant compartments, which are chest surface, lung surface, the atrial and ventricular myocardium and the cardiac blood masses. The ECG mapping data are acquired with a multi-channel ECG recording system [5,6].

In clinical practice the localization of ectopic origins is currently achieved by traditional catheter techniques and by recently introduced catheter mapping systems, like the CARTO^{TM} system [7]. The major advantage of the CARTO^{TM} system is that it enables the combination between electrical and spatial endocardial information, which supports the cardiologist in getting a deeper insight in the patients individual arrhythmia. CARTO^{TM} provides anatomical information that

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 588–595, 2006.

correlates to the electrogram mapping and the tagging system assists the creation of a linear lesion. CARTO^{TM} shows significant limitations when it comes to acquiring single-beat activation maps and, furthermore, these are mostly applied to single target chambers only. Another important fact is that it is impossible to obtain information from the epicardium under clinical conditions. Moreover, this technique is invasive which involves a risk to the patient during the procedure. These limitations are repealed and the information about the electrical excitation can be provided noninvasively, when applying the noninvasive imaging of cardiac electrophysiology (NICE) approach, provided that a VCM exists.

2 Implementation

For the semi-automatic segmentation of a VCM a framework has been developed using Java. The reason choosing Java was, that we have been developing web services for VCM modeling. A web service is a piece of logic that is placed on a server anyplace and can be operated using standard protocols like HTTP. The web service enables the functionality to handle distributed data, that exists when dealing with medical images in clinical settings, and allows to couple different services in order to generate new ones. Another important fact that argues for the usage of this technique is that the user only needs to have a Thin Client (usually an Internet Browser) installed, which should be standard on nearly every computer system, and the different time-consuming segmentation tasks can be executed parallel on different dedicated computing systems or a computing cluster.

The environment for executing the electrical field computation and for visualization of the models and results is amiraDev^{TM} 3.0 (TGS Europe Inc.). The amiraDev^{TM} environment is a visualization tool, which gives the opportunity to expand the functionality by implementing dynamic link libraries using the amiraDev^{TM} library.

3 Methods

3.1 Model Properties

When a VCM is used for imaging the electrical excitation in the heart a functional model instead of a detailed anatomical model is needed. The difference between an anatomical model and the desired functional VCM can be explained by the smoothing properties of the Fredholm integral, which describes the relationship between the source in the heart (transmembrane potential (TMP) distribution) and the potentials on the chest surface. Therefore, fine anatomical structures do not influence the macroscopic electrical properties, and can be smoothed out without decreasing the model's accuracy.

For the lung and chest extraction a T1 flash, non contrasted axial volume data set with a slice thickness of 10mm is taken. For the cardiac model, a cine-gated, T1-flash, non contrasted short-axis volume data set with a slice thickness of 6mm for the ventricles and 4mm for the atria is used.

3.2 Volume Conductor Model Extraction

The VCM used for estimating the electrical excitation in the patient individual heart consists of the compartments chest, lungs, atrial and ventricular myocardium, and the blood masses. A detailed explanation of the electrostatically most relevant compartments can be found in [8]. Because of the generally unknown individual fiber orientation, electrical isotropy is assumed in the surface heart model approach based on the bidomain theory.

Chest Extraction: The volume data are prefiltered using a Roberts Cross filter function [9], which highlights regions of high spatial gradient measurements which corresponds to edges. In the next step the minimal spanning bounding box (region of interest (ROI)) of the chest in the volume data is determined automatically. From the left and right upper voxel of the ROI a virtual beam with a 45 degree angle is projected to the chests border, which marks the border of the chest and tries to exclude the arms at those positions where they do not longer touch the body in order to get the torso.

Lung Extraction: The big advantage when trying to segment the lungs is, that the gray-values of the lungs differ intensely from the surrounding tissues. The main problem is, that the left and the right lungs are often in immediate contact, but for our electrical excitation estimation we need the lungs as separate compartments, due to implementation details of our inverse solver. Briefly the segmentation task can be described following way: first the volume data is thresholded to separate lungs from the surrounding fat, then the chest surrounding air, noise and airways are removed, and by usage of morphological operators the lung borders are rebuilt. In a last step the lungs are separated by a connected component labeling approach. More information about this approach can be found at [10].

Cardiac Blood Mass Extraction: Active Appearance Models (AAM) [11] have been shown to be useful for interpreting deformable objects. They contain statistical models of shape and gray-level appearance (texture) of objects and can be used for searching the desired structure in unseen images. The segmentation procedure is processed by synthesizing objects and by minimizing the error between the target and the synthesized object.

Objects in images are represented using shapes, which can be described by a set of n points. Shapes are placed in the training images by searching for corresponding landmark points, which was in our case done manually. A vector representing a shape can be defined as:

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

After having placed the landmark points of the different training images the shapes have to be realigned. Effects of scaling, translation and rotation have to be removed to enable statistics. All shapes are aligned in a way, that the sum of squared distances of each shape to the mean shape ($D = \sum |s_i - \overline{s}|^2$) is

minimized. As mentioned before, an AAM represents shape and texture changes learnt during the training. The shape can be represented as a vector s and the texture as a vector g:

$$s = \overline{s} + Q_s c, \tag{2}$$

$$g = \overline{g} + Q_g c, \tag{3}$$

where $\overline{s}$ represents the mean shape, $\overline{g}$ represents the mean texture, Q_s and Q_g are describing the modes of variation and c is the parameter vector controlling the shape and texture.

When using an AAM for segmentation the model seeks to minimize the difference between unseen image and one created by the appearance model. A difference vector δI can be defined as

$$\delta I = I_i - I_m, \tag{4}$$

where I_i describes the texture of the image and I_m the texture of the current model. The best match can be obtained when minimizing the magnitude of the squared difference vector:

$$\Delta = |\delta I|^2. \tag{5}$$

More information can be found in [12].

Ventricular Blood Mass Model: The models are generated using volume data from eight different cine-gated short-axis cardiac MRI scans. From each individual volume stack 15 images containing the ventricular blood masses at 4 different trigger time stages were taken. The landmarks were placed manually. For the left and the right ventricular blood mass boundary 33 landmarks were used, and more 33 landmark points were used for marking the outer boundary of the left ventricular myocardium.

Atrial Blood Mass Model: Due to the more complex geometry and, therefore, the more distinctive shape variations, the AAM model building approach needs to be more specific. To handle this task so called divisions of appearance are introduced. For the atrial blood mass extraction at least three divisions are necessary in order to cover the shape variations of the atria. The first division reaches from the base of the heart up to the left upper and the left lower pulmonary vein. The second division reaches from there up to the last appearance of the left upper and left lower pulmonary vein. From this position up to the right lower and right upper pulmonary vein the third division completes the overall AAM model. The same volume data of the eight patients was used to create the models. For each division 5 images at 4 different trigger states were taken. The mean Dice similarity coefficient (DSC) between the images taken for division one is 0.69, for division two 0.71, and for division three 0.67. The DSC between the images of the three divisions is below 0.56.

The landmarks were placed manually, and 33+33 landmark points were used to describe the right and left atrial blood mass shape. More 33 landmark points were used to define the cardiac boundary, which has the advantage that the initial model placing can be done more precisely.

Blood Mass Extraction Algorithm: Initially a few parameters need to be set. For the ventricles the apex cordis and the base of the heart needs to be marked and for the atrial blood mass segmentation the three divisions of appearance have to be marked in the image data. The segmentation approach can be described in pseudo code the following way:

1. *meanModel* = **get** mean model and calculate position of the model, where it has to be inserted initially
2. **forall** slides **do begin**
3. **insert** *meanModel* into the actual image slide
4. **AAM initialization** of the mean model:
   – find the blood masses without adopting the model the best possible way.
5. **solve minimization problem**; iteration until convergence (AAM fitting)
6. *meanModel* = model found during fitting procedure
   – (this enhances the segmentation speed and prevents from missplacing during AAM initialization, because the adjacent image is (usually) similar to the just segmented one)
7. **end forall**

Figure 1 shows the different steps of the AAM model fitting approach. In this sample the convergence was reached after 15 iterations.

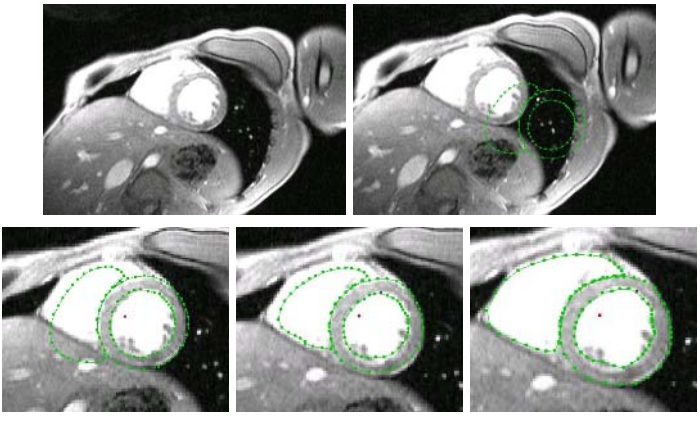


Fig. 1. Visualization of the AAM fitting approach. The shape models are superimposed to the medical images. The first image shows the medical image which is the input of the AAM segmentation task. The next slide shows the inserted mean model. The three images in the second row show the fitting process of the AAM search until convergence.

Myocardium Reconstruction: The myocardium reconstruction is based on a-priori knowledge provided by the atrial and ventricular blood masses. This approach employs mathematical and gray-value morphology operators [13]. It uses virtual circles as structuring elements with a radius range from a defined minimum up to a maximum wall thickness. The structuring element rolls

around the blood mass boundary and reconstructs the myocardial structure. If a voxel of the structuring element belongs to a gray-value range $g(V(x, y, z)) \in [minGray, maxGray]$, then the voxel is part of the myocardium. The above described case occurs when the myocardium can be detected in the volume data, but especially when reconstructing the atrial myocardium it happens that the myocardium can not be detected in the volume data by the algorithm. In this case anatomical mean model information is used to reconstruct or estimate the myocardial structure. As standard parameter 4mm are used for the atria and 8-10mm for the ventricles. More information can be found in [14].

Assembling the VCM: After having all relevant compartments segmented, it is necessary to generate the VCM. Therefore, all segmented labelsets were triangulated using the standard marching cubes algorithm amiraDev^{TM} provides. Because of the reason that the triangulation produces a lot of triangles an in-house developed amiraDev^{TM} plug-in is applied for remeshing and smoothing the generated surfaces. This process guarantees quality standards needed for solving the electrocardiographic inverse problem.

4 Results

Figure 2 shows three different patient individual VCMs. The left figure in row one shows a male, the right figure a female patient, where all compartments were extracted using the pipeline. The time for the extraction was about twelve minutes, and only a few manual corrections (atrial myocardium) had to be done. The second row shows a VCM of one patient in two different perspectives, without the compartment atria.

For testing the runtime behavior a 3.2Ghz CPU with 1 GB of main memory was used. Overall, in this procedure, ten patient data (different volume data compared to the AAM training sets) with structural normal hearts were taken to generate the VCMs. All patients underwent a electrophysiology study (EPS) because they suffered from Wolff-Parkinson-White syndrome. CARTO^{TM} maps, which are the gold-standard, were taken during the EPS study and compared with our model and our simulation results.

For the ventricular blood mass models one AAM was generated while the more complex geometry of the atria required to introduce three divisions of appearance and therefore, three AAMs were generated. This approach advanced the segmentation process for the atria.

The myocardium reconstruction is based of a-priori knowledge which is provided by the blood mass extraction method. In case the myocardium can be detected it can be reconstructed exactly, otherwise anatomical knowledge is used for reconstructing. This occurs predominantly when extracting the atrial myocardium.

The VCM integration was done by using the amiraDev^{TM} tool with in-house developed extensions.

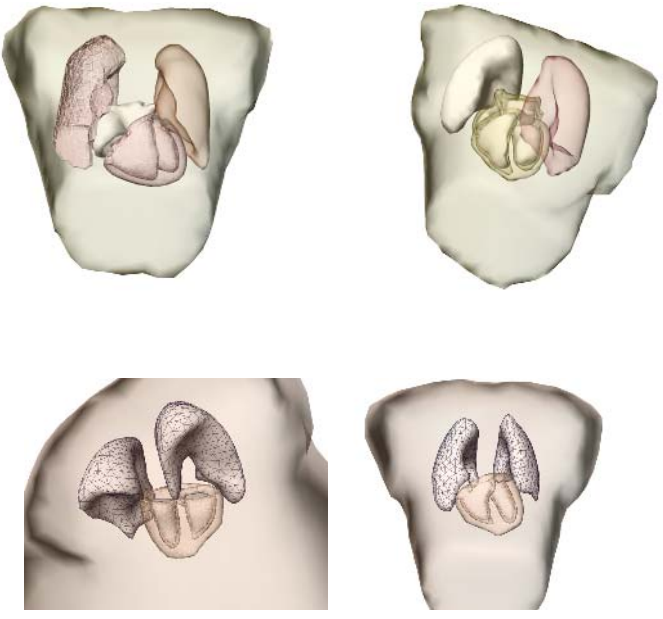


Fig. 2. Patient individual volume conductor models in different perspectives, visualized using amira^{TM}. For more details see text.

5 Discussion

The paper describes how a VCM can be extracted from morphological image data that qualifies for solving the electrocardiographic inverse problem.

One still existing problem is where the afferent blood masses (superior / inferior vena cava, pulmonary veins) should be separated. In our approach the afferent blood masses were separated as close as possible to the main target (the atria). This approach is supported by the fact that arrhythmias usually occur on the ventricle or atria and the veins have no or less electrical conductivity.

Furthermore, a major advantage of the proposed segmentation pipeline is, that only a few initial parameters have to be set and the overall runtime behavior is excellent. The generated models are of high quality and the model error is small, which is of importance when trying to solve the cardiographic inverse problem.

Because of the easy handling of the proposed approach, the good quality of the models and the good performance, this approach qualifies to be clinical applicable.

Acknowledgements

This study was supported by the Austrian Science Fund (FWF), grant START-Y144-N04.

References

1. Gulrajani, R.: Models of the electrical activity of the heart and computer simulation of the electrocardiogram. Crit. Rev. Biomed. Eng. **16**(1) (1988) p 1–66
2. Johnson, C.: Computational and numerical methods for bioelectric field problems. Crit. Rev. Biomed. Eng. **25**(1) (1997) p 1–81
3. Ramanathan, C., Ghanem, R., Jia, P., Ryu, K., Rudy, Y.: Noninvasive electrocardiographic imaging for cardiac electrophysiology and arrhythmia. Nat Med. **10**(4) (2004) p 422–8
4. Modre, R., Tilg, B., Fischer, G., Hanser, F., Messnarz, B., Schocke, M.S.M., Berger, T., Hintringer, F., Roithinger, F.: Atrial noninvasive activation mapping of paced rhythm data. J Cardiovasc Electrophysiol. 2003 Jul;14(7):720-1. **14**(7) (2003) p 720–9
5. van Rijn, R.A.C., Peper, A., Grimbergen, C.A.: High quality recording of bioelectric events. part 1 interference reduction, theory and practice. Med. Biol. Eng. Comput. **28**(5) (1990) p 389–397
6. van Rijn, R.A.C., Peper, A., Grimbergen, C.A.: Highquality recording of bioelectric events. part 2. low-noise, low-power multichannel amplifier design. Med. Biol. Eng. Comput. **29**(4) (1991) p 433–440
7. Andrea, E., Atie, J., Maciel, W., Araujo, N., Saad, E., Camanho, L., Affonso, H., Siqueira, L., Belo, L.G.: Mapping of supraventricular tachycardias by using a new three-dimensional technology: the carto system. J. Electrocardiol. **34**(4) (2001) p 334–334
8. Bradley, C., Pullan, A., Hunter, P.: Effects of material properties and geometry on electrocardiographic forward simulations. Annals of Biomedical Engineering **28**(7) (2000) p 721–741
9. Roberts, L.: Machine perception of 3-d solids. Optical and Electro-optical Information Processing. MIT press (1965) p 159–97
10. Heuberger, J., Geissbuehler, A., Mueller, H.: Lung CT segmentation for image retrieval. IPMI (2005)
11. Cootes, T., Taylor, C.: Active appearance models. 5th European Conference on Computer Vision; H. Burkhardt and B. Neumann (Springer 1998) p 484–498
12. Cootes, T., Taylor, C.: Active appearance models. http://www.isbe.man.ac.uk (2004)
13. Edward, R., Roberto, A.: Hands-on morphological image processing. SPIE Press, Bellingham, Washington USA (2003)
14. Pfeifer, B., Fischer, G., Hanser, F., Seger, M., Hintermueller, C., Modre-Osprian, R., Trieb, T., Tilg, B.: Atrial and ventricular myocardium extraction using model-based techniques. Methods Inf. Med. **45** (2006) p 19–26

Atrial Septal Defect Tracking in 3D Cardiac Ultrasound

Marius George Linguraru[1], Nikolay V.Vasilyev[2],
Pedro J. del Nido[2], and Robert D. Howe[1]

[1] Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA
[2] Dept. of Cardiac Surgery, Children's Hospital, Harvard Medical School, Boston, MA, USA
mglin@deas.harvard.edu

Abstract. We are working to develop beating-heart atrial septal defect (ASD) closure techniques using real-time 3-D ultrasound guidance. The major image processing challenges are the low image quality and the high frame rate. This paper presents comparative results for ASD tracking in sequences of 3D cardiac ultrasound. We introduce a block flow technique, which combines the velocity computation from optical flow for an entire block with template matching. Enforcing similarity constraints to both the previous and first frames ensures optimal and unique solutions. We compare the performance of the proposed algorithm with that of block matching and optical flow on six in-vivo 4D datasets acquired from porcine beating-heart procedures. Results show that our technique is more stable and has higher sensitivity than both optical flow and block matching in tracking ASDs. Computing velocity at the block level, our technique is much faster than optical flow and comparable in computation cost to block matching.

1 Introduction

Atrial septal defects (ASD) are congenital heart malformations consisting of openings in the septum between the atria. This allows blood to shunt from the left atrium into the right atrium, which decreases the efficiency of heart pumping. Although the surgical intervention for ASD closure is well established and has excellent prognosis, it is performed under cardiopulmonary bypass (CPB), which has widely acknowledged harmful effects. Minimally invasive beating-heart ASD closure will avoid using CPB and improve the patient rehabilitation. Reliable visualization of structures within the heart has been one of the major challenges to successful beating-heart surgical interventions [12].

3D Ultrasound (US) is simple, cheap and fast, and allows the surgeon to visualize cardiac structures and instruments through the blood pool. US also has some major disadvantages, being extremely noisy with poor shape definition, which makes it confusing and hard to interpret in the operation room. Tracking tissue in 3D US volumes is particularly difficult due to the low spatial resolution. Therefore, the development of tracking methods for volumetric data in 4D applications is necessary to assist clinical procedures.

An optical flow approach was proposed in [5] to track endocardial surfaces. The data points are initialized manually and a finite element surface is fitted to the points.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 596 – 603, 2006.

The method extends to 3D the region-based optical flow from [11] using a simple correlation and takes approximately 2 min/frame after initialization. Boukerroui *et al.* [2] use a better-adapted similarity measure introduced in [4] to compute region-based optical flow in US image sequences. Their results, although faster, are 2D. The initialization is left out as a separate subject and the authors propose a parameter optimization scheme [2]. A knowledge-based approach using level sets is proposed in [9]. Segmentation and tracking are alternated using shape knowledge, visual information and internal smoothness constraints. The method is parametric and 2D. An interesting 3D cross-correlation-based approach for speckle tracking on simulated data is proposed in [3]. However, 3D speckle tracking poses a serious of difficulties, from optimization and computational costs, to speckle decorrelation in time and space.

A number of approaches have been tried. Block matching is fast and simple, but estimates velocities at low level and lacks robustness. Optical flow has higher sensitivity and specificity, but is very slow and must find a good compromise between local and global displacements.

In this paper, we present an approach adapted to ASD tracking that estimates velocities for an entire block using the concept of voxel-based optical flow. The method is fast and avoids the problems of the traditional block matching, while using the sensitivity of optical flow. The computation is done volumetrically and the displacement optimizes a similarity measure relative to both the previous and original frames. We compare the new results to those obtained by classical implementations of block matching and optical flow.

Section 2 of the paper presents the methodology of the tracking algorithms. We mention block matching, optical flow and similarity measures, and introduce the block flow technique. Comparative results on 3D cardiac US images are shown in Section 3.

2 Method

At this stage, the initialization of the process is done manually by allowing the user to select a block in the first frame that is centered on the ASD. This gives a first estimate of the ASD location and a template for the computation of similarity scores, as seen later in the paper.

2.1 Block Matching

The first approach we considered was that of single block matching [8]. This multi-scale intensity-based method assumes that there is a global intensity relationship between two images. We are simply interested in matching two blocks, which simplifies the algorithm. We investigated three types of similarity measure suitable for images of the same modality. First, an identity measure, the sum of squared differences (*SSD*) of normalized intensities, as seen in Equation (1); a linear measure, the inverse Pearson correlation coefficient (*PCC*), as in Equation (2); and a statistical

measure, the maximum likelihood (*MLE*) introduced in [4] and refined in [2], as seen in Equation (3). *ref* is the reference block, *tar* the target and *n* represents the number of voxels in a block. An overview of intensity similarity measures can be found in [10]. While *SSD* does not cope with relative intensity, *PCC* minimizes the least square fitting to the original data. The *MLE* measure [2] assumes that US images are log compressed with Rayleigh noise, but no change of noise between frames. Minimizing the square error between blocks becomes a matter of minimizing the probability distribution function of the additive noise between frames.

$$\overline{ref} = \left(ref - \mu_{ref}\right)/\sigma_{ref};\ \overline{tar} = \left(tar - \mu_{tar}\right)/\sigma_{tar};$$
$$SSD = \frac{1}{n}\sum_i \left(\overline{ref}_i - \overline{tar}_i\right)^2; \tag{1}$$

$$PCC = \frac{1}{n\sum_i \overline{ref}_i * \overline{tar}_i}; \tag{2}$$

$$MLE = \frac{1}{n}\sum_i \left(\ln \overline{ref}_i - \ln \overline{tar}_i - \ln\left(\exp\left(2\left(\ln \overline{ref}_i - \ln \overline{tar}_i\right)\right)+1\right)\right); \tag{3}$$

Our implementation uses two steps, from coarse to fine, and finds the block in the target image that is most similar to the given block in the reference image. To avoid padding at margins, we reduce the size of the reference block to account for clipping at the edge of the target. The best-matched target block is then grown to the original size of the reference.

2.2 Optical Flow

As a second approach, we tested the efficiency of region-based optical flow. This is a 3D extension of the 2D method presented in [2] and based on [11]. There are other techniques to compute optical flow, such as the first-order differential method in [7]. Such methods are very sensitive to noise, which is high in US, and require spatial and temporal smoothing. Our real-time tracking application cannot smooth temporally, as we do not know the data in the next frames. Phase-based techniques [6] require an optimal search space, which is difficult to predict in cardiac images. A review of optical flow techniques can be found in [1].

The major disadvantage of optical flow techniques is the computational load. We propose a two-step multi-scale approach, which uses both block matching and optical flow. Block matching gives the coarse velocity, which is subsequently refined by optical flow. At both levels, we employed *MLE* as similarity measure. First, the contour of the ASD is extracted in the first frame and then the velocity of the entire block is estimated. At the finer level, the optical flow is estimated only for the contour voxels, which are then wrapped by the resulting block. This approach is considerably faster than multi-scale optical flow.

2.3 Block Flow

Finally, we introduce the notion of "block flow", which uses the velocity estimation of optical flow [11], but for the entire block, instead of every voxel. For a given reference block *ref*, we define empirically a maximum displacement *md* in the new image frame. Within the search space, we compute the estimation error as in Equation (4). Minimizing the energy E is equivalent to minimizing the maximum error in similarity between the target block *tar* and *ref*, and between *tar* and *absref*, where *absref* is the user-defined block in the first frame of the 4D sequence. E becomes a value of worst match and reduces velocities u, v and w in directions of blocks that are not similar to both the previous frame and the absolute reference. *absref* acts as a template ensuring that the solution does not become too different than the initial estimate.

Extending to 3D the principles presented in [2, 11], we obtain the probability distribution R, as in Equation (5), where τ normalizes the probabilities to sum to 1. The estimate of the three-directional block velocity V_b is presented in Equation (6). Singh [11] minimizes its error through a Gauss-Siedel relaxation algorithm. Unlike [11], we do not use neighborhood information, since the displacement is unique for the entire block. As this enforcement is valid for tracking an object like ASD, this is not applicable to any tracking problem.

$$E(u,v,w) = \max\left(MLE\left(\overline{ref}, \overline{tar}\right), MLE\left(\overline{absref}, \overline{tar}\right)\right); \tag{4}$$

$$R(u,v,w) = \frac{1}{\tau}\left(\exp\left(-\frac{E(u,v,w) - \max\left(E(u,v,w)\right)}{2md^3}\right) - 1\right); \tag{5}$$

$$V_b = \left(\sum_{u,v,w} R(u,v,w)u, \sum_{u,v,w} R(u,v,w)v, \sum_{u,v,w} R(u,v,w)w\right); \tag{6}$$

Once the velocity is calculated, we verify if the new match is on the image margins. If that is the case, we grow the block to ensure that it will propagate with the same size as *absref*.

3 Results

To test the tracking algorithms, we used a database of six 4D time sequences of porcine beating hearts with artificially created ASD. All the US data were acquires in-vivo with a Sonos 7500 Live 3D Echo scanner (Philips Medical Systems, Andover, MA, USA). The time of acquisition was of 2s/case at a frame rate of 25 volumes/s. At this stage, the implementation of the three tracking algorithms was done in Matlab version 7 (MathWorks, Inc.) on a Pentium IV machine with 1GB RAM and 2.40 GHz processor. The image size is 80x80x176x50.

The performance of the block matching algorithm using two levels (coarse to fine) of combined translation-rotation and *MLE* is shown on the top row of Figure 1. Each block shown in Figure 1 is a 3D entity visualized from the right atrium looking into

the left atrium. The rows present the absolute reference in frame 1 and tracking results after 20, 40 and 50 frames. Judging by the small rotations and the rounder shape of ASD, we also tested block matching using only multi-scale translation. The results are almost identical (with an error of maximum 2 voxels) and the speed is increased to 2 frames/s with *SDD* and *PCC* similarity criteria. Using *MLE* as similarity measure gives more sensitivity to the algorithm, but it is more computationally expensive (0.5 frame/s).

The performance of the two level optical flow algorithm (block matching at the coarse level and optical flow at the fine level) was visually more robust than that of simple block matching. Using *MLE* as similarity criterion, the computational speed was of 1 min/frame. The middle row in Figure 1 shows tracking results using optical flow.

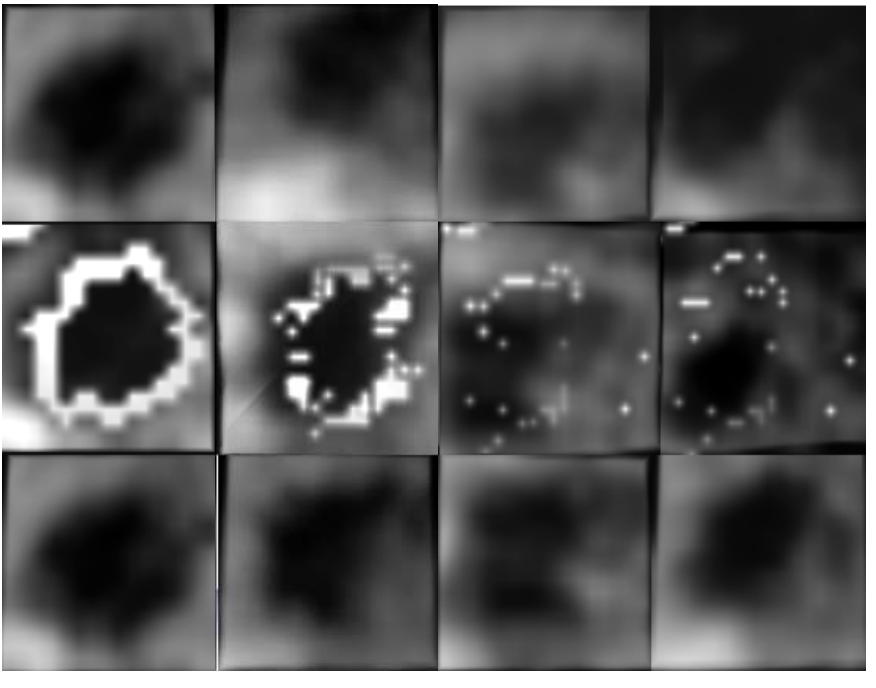


Fig. 1. The comparative performance of the three tracking algorithms. The top row shows results by block matching; the middle row by optical flow; the bottom row by block flow. All three examples used *MLE* as similarity criterion. From left to right, the columns show 3D blocks visualized from right atrium to left atrium at frames 1, 20, 40 and 50. In the middle row we also show the migration of ASD contour voxels (bright spots), as their velocities are computed by optical flow. While we used smoothing between neighboring points, no connectivity was enforced.

Finally, the bottom row in Figure 1 presents tracking results using our block flow algorithm. Using energy estimation at block level instead of voxel level, the block

flow method is much faster than optical flow. Its computation speed is 0.3 frame/s, comparable to that of simple block matching, but with much more robust results.

In Figure 2, we present larger 3D volumes (the entire frame) of the results shown in Figure 1. Note that the migration of contour voxels in the optical flow algorithm enlarges the ASD block. Although in Figure 1 the performance of optical flow and block flow look similar and better than that of block matching, Figure 2 shows superior results of our block flow technique.

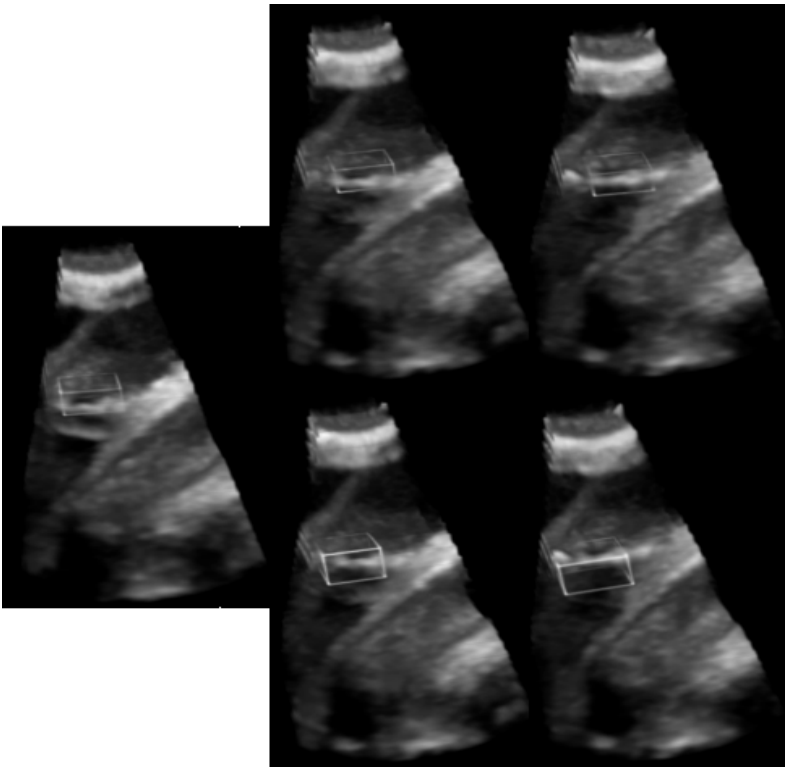


Fig. 2. Comparative tracking results in the full 3D frame. The left column presents the first frame and the absolute reference marked as a block; the middle column shows tracking results after 25 frames; the right column after 50 frames. The top row shows results by block flow and the bottom row by optical flow. Please note the enlarged blocks on the bottom row.

Table 1 shows tracking quantitative results. Mean *MLE* scores are computed for the six 4D images using the three tracking algorithms. The minimum score represents the best match. The block flow algorithm performed better than the other two, with the exception of one case, where the optical flow technique had lower scores. Block matching performs better than optical flow over the first 25 frames, due to the error induced by the block growth, but is less stable in time.

Table 1. The mean *MLE* scores for the six ASD 4D images using the three tracking techniques. Scores were calculated after 25 and 50 frames as similarities with the absolute reference. *MLE* scores are normalized and vary from 0 for the perfect match to 100 for the worst. The block flow algorithm performs better in all but one case, where tracking with optical flow has a lower score.

	Frame 25	Frame 50
Block Matching	6.82	9.29
Optical Flow	7.46	7.17
Block Flow	4.14	5.99

4 Discussion

We presented comparative results for the tracking of ASD in sequences of 3D cardiac ultrasound. We introduced a block flow technique, which combined the velocity computation from optical flow with template matching. Enforcing similarity constraints to both the previous and first frames, we ensure optimal and unique solutions. Computing velocity at the block level, our technique is much faster than optical flow and comparable in computation to block matching. Results on in-vivo 4D datasets show that our technique is more stable and has higher sensitivity than both optical flow and block matching in tracking ASD.

The block flow tracking algorithm finds the three-dimensional velocity of an entire block that contains the object of interest. In this application, the object is an ASD with an approximate cylindrical shape. The ASD can be extracted from every new block by simple and fast thresholding.

Compared to the basic block matching algorithm, out method has higher sensitivity in computing displacements, which gives it more robustness in time. Optical flow is sensitive to very localized changes, which makes the method also prone to errors. Interpolation, smoothing and connectivity constraints improve the functioning of optical flow, but also approximate the results. Moreover, the block flow method is much faster than optical flow. For an application such as ASD tracking, the block displacement is sufficient, but this will be too coarse for tracking finer details.

We first normalized the intensities to reduce the effect of noise change and angular reflections from frame to frame. Then we compared the performance of several similarity measures and found that statistical measures, such as maximum likelihood, are better for our application. A coarse to fine approach is desirable for effectiveness and speed.

We used a Rayleigh noise model in the similarity measure. Although we did not smooth the data, the pre-processing of the commercial ultrasound machines may alter the noise distribution. Our assumption led to robust results, but other noise models may be studied.

For future work, we will optimize the algorithm implementation to make it run in real time. The automatic detection of the ASD absolute reference will be investigated along with the use of the cardiac cycle for predictive estimation.

Acknowledgements

This work is funded by the National Institutes of Health under grant NIH R01 HL073647-01. The authors would like to thank Dr. Gerald Marx from Children's Hospital in Boston, and Dr. Ivan Salgo from Philips Medical Systems for assistance with image acquisition and informative discussions.

References

[1] J.L. Barron, D.J. Fleet, S. Beauchemin: Performance of Optical Flow Techniques. In: International Journal of Computer Vision, Vol. 12(1) (1994) 43-77.

[2] D. Boukerroui, J.A. Noble, M. Brady: Velocity Estimation in Ultrasound Images: a Block Matching Approach. In: Information Processing in Medical Imaging (IPMI) (2003) 586-598

[3] X. Chen, H. Xie, R. Erkamp, K. Kim, C. Jia, J.M. Rubin, M. O'Donnell: 3-D Correlation-based Speckle Tracking. In: Ultrasonic Imaging. Vol. 27(1) (2005) 21-36

[4] B. Cohen, I. Dinstein: New Maximum Likelihood Motion Estimation Schemes for Noisy Ultrasound Images. In: Pattern Recognition, Vol. 35 (2002) 455-463

[5] Q. Duan, E.D. Angelini, S.L. Herz, C.M. Ingrassia, O. Gerard, K.D. Costa, J.W. Holmes, A.F. Laine: Evaluation of Optical Flow Algorithms for Tracking Endocardial Surfaces on Three-dimensional Ultrasound Data. In: SPIE International Symposium, Medical Imaging Vol. 5750 (2005)

[6] D.J. Fleet, A.D. Jepson: Computation of component image velocity from local phase information. In: International Journal of Computer Vision, Vol. 5(1) (1990) 77-104

[7] B. Lucas, T. Kanade: An Iterative Image Registration Technique with an Application to Stereo Vision. In: Proc. Image Understanding Workshop (1981) 121-130

[8] S. Ourselin, A. Roche, S. Prima, N. Ayache: Block Matching: A General Framework to Improve Robustness of Rigid Registration of Medical Images. In: A.M. DiGioia and S. Delp (eds.): Medical Image Computing and Computer Assisted Intervention (MICCAI 2000), Lectures Notes in Computer Science Vol. 1935 (2000) 557-566

[9] N. Paragios: A Level Set Approach for Shape-driven Segmentation and Tracking of the Left Ventricle. In: IEEE Trans Med Imaging, Vol. 22(6) (2003) 773-786

[10] A. Roche, G. Malandain, N. Ayache: Unifying Maximum Likelihood Approaches in Medical Image Registration. In: International Journal of Imaging Systems and Technology: Special issue on 3D imaging, Vol. 11 (2000) 71-80

[11] A. Singh, P. Allen: Image-flow Computation: an Estimation-theoretic Framework and a Unified Perspective. In: CVGIP: Image Understanding, Vol. 65(2) (1992) 152-177

[12] Y. Suematsu, G.R. Marx, J.A. Stoll, P.E. DuPont, R.O. Cleveland, R.D. Howe, J.K. Triedman, T. Mihaljevic, B.N. Mora, B.J. Savord, I.S. Salgo, P.J. del Nido: Three-Dimensional Echocardiography-guided Beating-heart Surgery without Cardiopulmonary Bypass: a Feasibility Study. In: Journal of Thoracic and Cardiovascular Surgery, Vol. 128(4) (2204) 579-587

Intra-operative Volume Imaging of the Left Atrium and Pulmonary Veins with Rotational X-Ray Angiography

Robert Manzke[1], Vivek Y. Reddy[2], Sandeep Dalal[1],
Annemarie Hanekamp[1], Volker Rasche[3], and Raymond C. Chan[1]

[1] Philips Research North America, Clinical Sites Research,
345 Scarborough Road, Briarcliff Manor, NY, USA
robert.manzke@philips.com
[2] Experimental Electrophysiology Laboratory, Cardiac Arrhythmia Service,
Massachusetts General Hospital, Boston, USA
[3] University Clinic Ulm, Germany

Abstract. Complex electrophysiology (EP) procedures, such as catheter-based ablation in the left atrium and pulmonary veins (LAPV) for treatment of atrial fibrillation, require knowledge of heart chamber anatomy. Electroanatomical mapping (EAM) is typically used to define cardiac structures by combining electromagnetic spatial catheter localization with surface models which interpolate the anatomy between EAM point locations in 3D. Recently, the incorporation of pre-operative volumetric CT or MR data sets has allowed for more detailed maps of LAPV anatomy to be used intra-operatively. Preoperative data sets are however a rough guide since they can be acquired several days to weeks prior to EP intervention. Due to positional and physiological changes, the intra-operative cardiac anatomy can be different from that depicted in the pre-operative data.

We present a novel application of contrast-enhanced rotational X-ray imaging for CT-like reconstruction of 3D LAPV anatomy during the intervention itself. We perform two selective contrast-enhanced rotational acquisitions and reconstruct CT-like volumes with 3D filtered back projection. Two volumes depicting the left and right portions of the LAPV are registered and fused. The combined data sets are then visualized and segmented intra-procedurally to provide anatomical data and surface models for intervention guidance. Our results from animal and human experiments indicate that the anatomical information from intra-operative CT-like reconstructions compares favorably with pre-acquired CT data and can be of sufficient quality for intra-operative guidance.

1 Introduction

Currently, a large number of interventional cardiac electrophysiology (EP) procedures are performed with fluoroscopic X-ray guidance alone. Two-dimensional projections from a single view are typically used to visualize catheter location and EP device placement. Minimally-invasive catheter-based ablation has become the first line therapy for many cardiac arrhythmias. However, attempts to ablate more complex arrythmia, such as atrial fibrillation (AFib), continues to be very challenging [1] because of limitations in projection imaging. Fluoroscopy times longer than one hour are not uncommon for such complex procedures, resulting in significant X-ray exposure to the patient and

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 604–611, 2006.

clinical staff [2, 1]. Ablation strategies based on anatomical information from electroanatomic mapping (EAM) can improve the efficacy of catheter ablation for these complex cases [2, 1, 3], reducing the X-ray dose significantly.

EAM mapping can however be a tedious procedure which yields a sparse number of cardiac surface points through which an interpolating surface is lofted as an aid to visualization. Recently, pre-operative volumetric imaging techniques such as magnetic resonance imaging (MRI) and computed tomography (CT) have been used to provide more detailed representations of the LAPV anatomy in 3D. However, due to factors including patient positioning, volume status, and physiological variation over the often long time interval between pre-operative imaging and intervention, the anatomy from pre-operative data can be quite different from the intra-procedural heart chamber shape. MRI cannot currently be used for patients with device implants that are not MR-compatible; this clinical population represents a sizeable fraction of patients undergoing EP procedures. CT on the other hand increases the X-ray radiation exposure for the patient, beyond the dose already necessary in the intervention. Despite these limitations, the level of anatomical detail provided by pre-operative data sets has significantly improved LAPV ablation guidance when integrated intra-operatively with catheter-tracking-based EAM systems [3, 4].

To address the limitations of pre-operative imaging, we have developed a novel approach for intra-procedural 3D imaging of the LAPV anatomy. With contrast-enhanced 3D rotational X-ray angiography (3DRA) in combination with 3D filtered back projection algorithms, we are able to reconstruct the anatomy of the LA and the PVs in a CT-like fashion. We use selective contrast injections to enhance separate rotational acquisitions of the left and right portions of the LAPV structure. Following CT-like reconstruction, the two volumes are registered and fused. We then use the volumetric data intra-operatively for direct volume or cut-plane visualization and for LAPV surface segmentation.

This paper is structured as follows: Section 2 describes our methods and materials. Section 3 presents our results from CT-like 3DRA reconstruction relative to pre-operatively acquired CT data sets. Our conclusions and future work are given in section 4.

2 Methods and Materials

2.1 Clinical Imaging Protocols

3D Rotational X-Ray (3DRA). 3DRA data acquisition is performed using the Philips Allura XPer FD10 flat detector single-plane X-ray system. Two breath-hold acquisitions are performed, each with selective iodine contrast agent injection of 60cc at a rate of 20cc/s into the left/right branch of the pulmonary arteries.

The transit time of each contrast bolus through the pulmonary circulation and into the LAPV anatomy is somewhat patient-dependent and takes approximately 3-4s after initiating the injection (see Figure 1). Rotational acquisition is initiated, accounting for this delay interval, and takes 4s over which the C-arm rotates 220 degrees around the patient. 120 projections are captured at 30Hz frame rate. The overall time for each acquisition requires $\sim$10s, a very reasonable breath-hold duration.

Multislice Computed Tomography (MSCT). For comparison of 3DRA with pre-operatively acquired MSCT data sets, imaging was performed using a standard cardiac helical imaging protocol (120 kV, 500 mAs, 64×0.75mm collimation, 0.42 s rotation time) together with a conventional intravenous contrast injection protocol (20cc timing bolus, and imaging bolus of 100cc at 5cc/s). ECG-gated images were reconstructed with 0.75mm slice thickness, at 0.3mm increments, with a 20cm field-of-view, a medium smooth reconstruction filter, and a 512x512 pixel matrix.

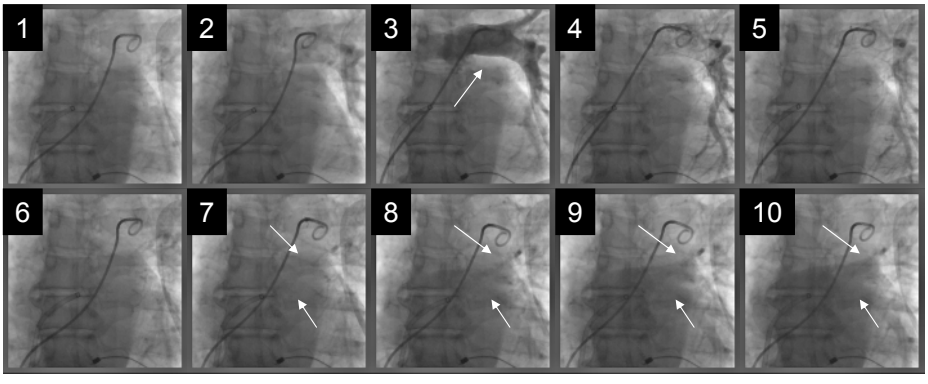


Fig. 1. Cine acquisition of a bolus injection into the left pulmonary artery (20cc at 20cc/s) where each frame corresponds to 1s. Test bolus injection takes a total of 1s and the contrast agent passes the lung in about 4s. Arrows mark opacified left PV's.

2.2 Image Reconstruction

Cone-beam projections, $p(\beta, \mu, \nu)$, from 3DRA were acquired, where β is the rotational angle, and (μ,ν) are the detector coordinates. Un-gated volume reconstruction was then performed with the standard approximate Feldkamp (FDK) cone-beam reconstruction algorithm [5], which is available on the Philips Allura platform:

$$w(\mu, \nu) = \frac{\overline{SO}}{\sqrt{\overline{SO}^2 + \mu^2 + \nu^2}}\,, \tag{1}$$

$$p^f(\beta, \mu, \nu) = \{w(\mu, \nu) \cdot p(\beta, \mu, \nu)\} * h(\mu)\,, \tag{2}$$

$$f(\mathbf{x}) = \int_0^{2\pi} \frac{\overline{SO}^2}{U^2(x, y, \beta)} p^f(\beta, \mu(x, y, \beta), \nu(\mathbf{x}, \beta)) d\beta\,, \tag{3}$$

In these equations, the distance from source to the detector is $\overline{SO}$, $h(\cdot)$ is the ramp filter kernel, $\mathbf{x} = (x, y, z)^T$ is the voxel coordinate and $f(\cdot)$ is the object function. $w(\mu, \nu)$ specifies the cone-angle dependent pre-weighting and $U(\cdot)$ is the voxel-dependent weighting of the FDK algorithm. Geometry calibration of the X-ray system is performed prior to acquisition [6].

2.3 Registration, Fusion, and Segmentation of 3DRA-Derived Volumes

Following volumetric reconstruction of data sets from the left and right pulmonary artery injections, we registered the two volumes using the thoracic vertebrae present in both acquisitions. Slight variations in breath-holds during the acquisition were disregarded. A sub-volume of interest large enough to contain the spine in both data sets was graphically-prescribed. Starting with an initial estimate of the transformation between the left and right sub-volumes in addition to trilinear interpolation of voxel intensities, the right volume was transformed into the coordinate-space of the left volume. The resulting sub-volumes were vectorized using lexicographical-ordering and normalized. The 3D normalized correlation (Equation 4) was then computed between the data sets as follows:

$$C = \hat{I}_L^T(\mathbf{x}_L)\hat{I}_R(\mathbf{T}(\mathbf{x}_R)) \tag{4}$$

where $\hat{I}_L(\mathbf{x}_L)$ and $\hat{I}_R(\mathbf{T}(\mathbf{x}_R))$ are respectively the left and right normalized volume vectors, $\mathbf{x}_L$ designates left volume coordinates, $\mathbf{x}_R$ designates right volume coordinates, and $\mathbf{T}(\mathbf{x}_R)$ designates right volume coordinates transformed into the left volume coordinate system. The cross-correlation objective function was iteratively maximized using a constrained-search of the transformation parameter space. Afterwards, the two volumes were fused for visualization and further analysis by using the average volume intensity at each grid point in the combined data set. The mean intensities of the individual volumes, $\bar{I}_L$, and $\bar{I}_R$, were used to normalize each voxel intensity before averaging as shown in Equation 5.

$$I(\mathbf{x}) = \frac{1}{2}\left\{\frac{\bar{I}_L}{\bar{I}_R} \cdot I_R(\mathbf{x}) + I_L(\mathbf{x})\right\} \tag{5}$$

Figure 2 demonstrates a result from spine registration and fusion from a sample data set.

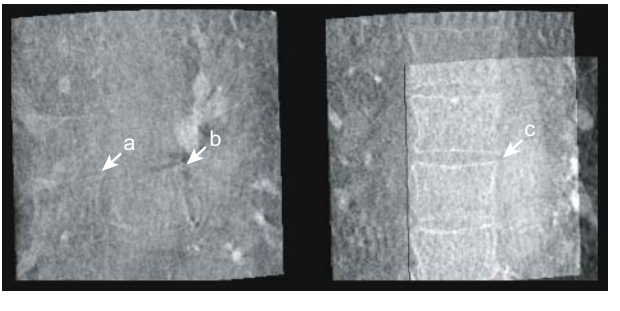


Fig. 2. Sample result to illustrate volumetric spine fusion with mis-registration on the left, and accurate registration on the right. Annotations (a) and (b) show two separate sets of spinal vertebrae due to lack of registration. Annotation (c) in the registered volume clearly shows a single vertebral column, reflecting satifactory registration.

We segmented the fused volume by graphically-specifying a seed point within the contrast-enhanced LAPV structure. With prescribed limits for the minimum and maximum voxel intensities within the LAPV, we performed threshold-based region-growing around the initial seed point to segment in 3D all connected voxels falling within the LAPV region-of-interest. Morphological erosion and dilation were used to eliminate spurious edge voxels and holes in the segmentation. A triangulated surface model was extracted from the segmented binary mask using the marching cubes isosurfacing algorithm [7] and further smoothed for comparison with models extracted from pre-operatively acquired MSCT data.

3 Results

3.1 Animal Results

Figure 3 shows volume rendered images of a reconstructed heart from porcine imaging. The procedure described in Section 2.1 was used for contrast injection, except that the contrast agent was injected into the main branch of the pulmonary artery. Due to the higher heart rate and smaller stroke volume for pigs, the contrast bolus enhanced the aorta and the left ventricle, in addition to the LAPV. The small size of the pig heart allowed for complete volume coverage in a single acquisition, obviating the need for volume registration. Good anatomical correlation between the reconstructions from 3DRA (upper row, Figure 3) and MSCT (lower row) can be appreciated. In particular, the LAPV structure is clearly identifiable within the reconstructed 3DRA data set.

3.2 Human Results

Acquisitions with selective enhancement of the left and right sides of the LAPV structure were performed (see Figure 4). Separate acquisitions were required due to field-of-view limitations of the X-ray system and rapid contrast dilution as the bolus propagates through the cardiovascular system.

A qualitative comparison of the CT-like 3DRA data relative to pre-operative CT imaging is shown in Figure 5. As it was observed for the porcine imaging study, the LAPV anatomy can clearly be appreciated. Relative to MSCT, the rotational angiography reconstructions are noisier, due to the lower angular sampling of projection data and lower detector dynamic range. The level of structural detail within the CT-like 3DRA reconstruction (top row of Figure 5) is comparable to that present within the reference MSCT volume despite the fact that reconstructions from 3DRA were performed without any cardiac gating. This finding is attributable to the low degree of LAPV motion over the cardiac cycle, especially in the presence of atrial fibrillation which disrupts the synchrony of LAPV motion.

In the last Figure 6, a segmentation surface model of the human LAPV structure is shown. Such a surface model can be used in combination with EAM to guide mapping and intervention. Segmented surfaces from 3DRA (left image) show anatomical features within the LAPV that are similar to those present in the segmented MSCT data. Segmentation of CT-like 3DRA is more challenging than MSCT data due to the poorer

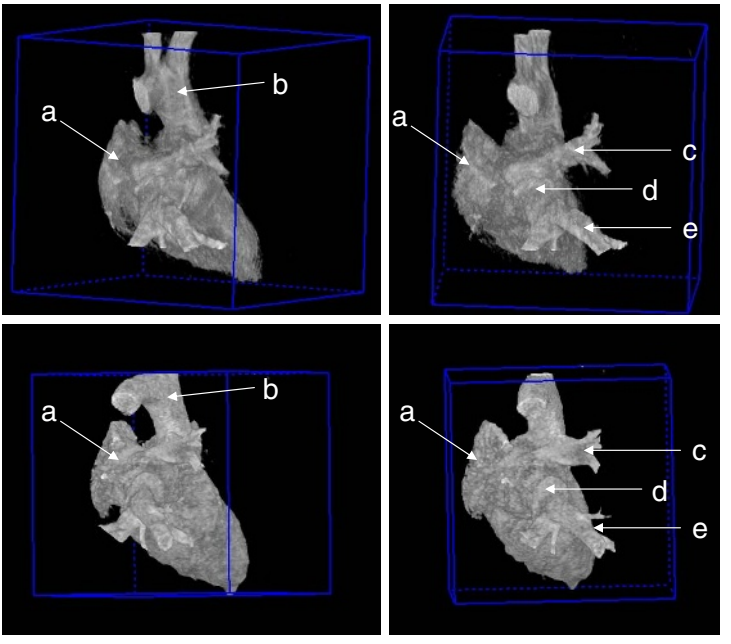


Fig. 3. Upper row shows volume renderings of a CT-like 3DRA reconstruction from a porcine study (posterior view orientation). (a) depicts the left atrial appendage, (b) is the aortic arc, (c) shows the right superior PV, (d) shows the LA and (e) indicates the right inferior PV. Lower row displays renderings from pre-operative MSCT imaging.

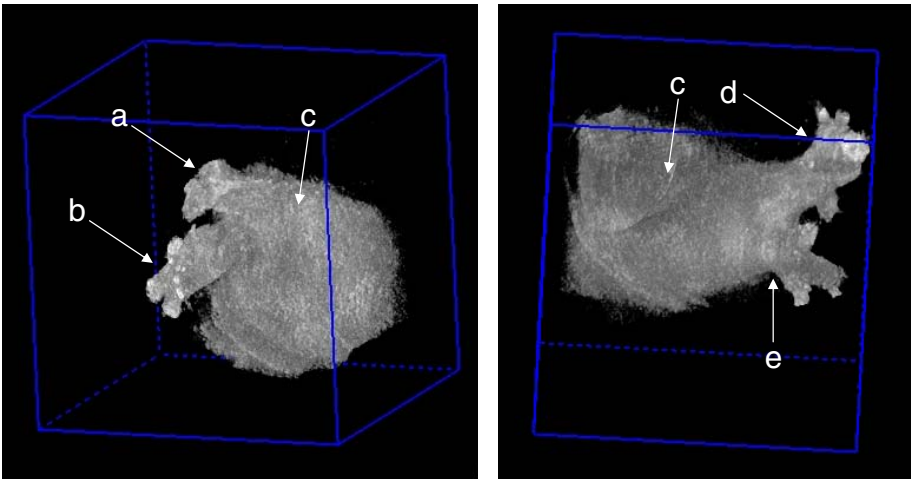


Fig. 4. On the left, a posterior volume rendering of the selective left pulmonary artery injection is shown. (a) reflects the left superior PV, (b) points at the left inferior PV and (c) shows a part of the LA. Right image illustrates the selective right pulmonary artery injection. (c) shows a part of the LA, (d) shows the right superior PV and (e) shows the right inferior PV.

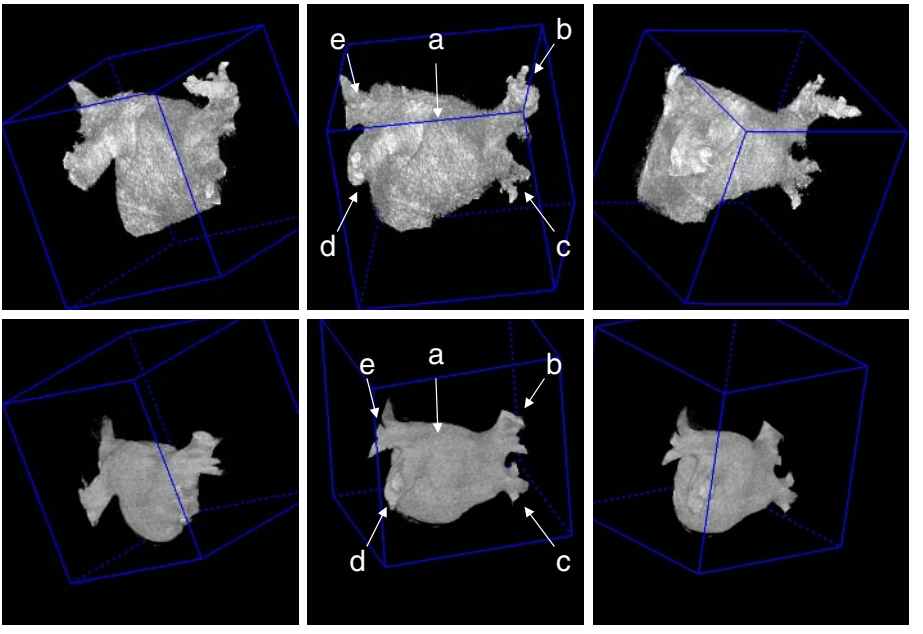


Fig. 5. Top row shows volume renderings from the reconstructed 3DRA run. (a) indicates the LA structure, (b) shows the right superior PV, (c) depicts the right inferior PV, (d) shows the left inferior PV and (e) indicates the left superior PV. Lower row displays renderings from the corresponding pre-operative CT data.

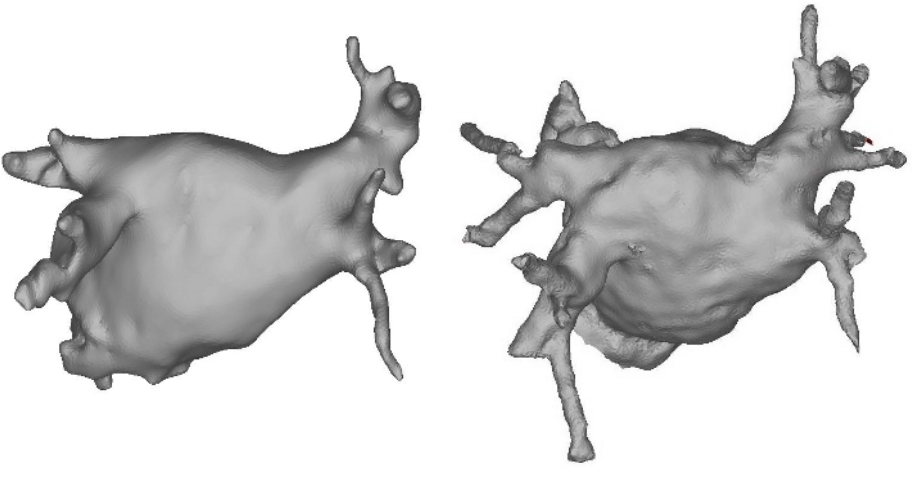


Fig. 6. Left image shows a view of the LAPV surface model, segmented from the fused 3DRA data set. Right image shows a segmentation from MSCT.

signal-to-noise characteristics, contrast levels, and artifacts from catheters and motion within the imaging field.

4 Conclusion and Future Work

We have presented a method for intra-operative volumetric imaging data of the LAPV anatomy that reflects the patient's anatomy at the moment of the intervention. This technique has the potential to be used during clinical AFib ablation procedures for improved 3D interventional navigation and may eliminate the need for pre-interventional volume imaging. Comparisons with pre-operative CT data reveal good qualitative correspondence. Compared with MSCT, a smaller X-ray dose is required to obtain useful images of the LAPV anatomy, reducing the overall radiation exposure to the patient.

Our future work will focus on a quantitative comparison with pre-operative volumetric images. Further optimization for automated acquisition protocol timing to account for pulmonary transit times will be performed. Finally, improvements in segmentation tools for automated surface model extraction will be explored.

References

[1] Haissaguerre, M., Jais, P., Shah, D.C., Takahashi, A., Hocini, M., Quiniou, G., Garrigue, S., Mouroux, A.L., Metayer, P.L., Clementy, J.: Spontaneous initiation of atrial fibrillation by ectopic beats originating in the pulmonary veins. New England Journal of Medicine **339** (1998) 659 – 666

[2] Pappone, C., Oreto, G., Rosanio, S., et al.: Atrial electroanatomic remodeling after circumferential radiogrequency pulmonary vein ablation: efficacy of an anatomic approach in a large cohort of patients with atrial fibrilation. Circulation **104** (2001) 2539 – 2544

[3] Reddy, V.Y., Malchano, Z.J., Holmvang, G., Schmidt, E.J., dAvila, A., Houghtaling, C., Chan, R.C., Ruskin, J.N.: Integration of cardiac magnetic resonance imaging with three-dimensional electroanatomic mapping to guide left ventricular catheter manipulation. Journal of the American College of Cardiology **44**(1) (2004) 2202 – 2213

[4] Mikaelian, B.J., Malchano, Z.J., Neuzil, P., Weichet, J., Doshi, S.K., Ruskin, J.N., Reddy, V.Y.: Integration of three-dimensional cardiac computed tomography images with real-time electroanatomic mapping to guide catheter ablation of atrial fibrillation. Circulation **112** (2005) 35 – 36

[5] Feldkamp, L.A., Davis, L.C., Kress, J.W.: Practical cone-beam algorithm. Journal of the Optical Society of America **A6** (1984) 612 – 619

[6] Movassaghi, B.: Ph.D. thesis on 3D rotational X-ray coronary angiography. Utrecht University, ISBN:90-393-05048 (2004)

[7] Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3d surface construction algorithm. In: SIGGRAPH '87: Proceedings of the 14th annual conference on Computer graphics and interactive techniques, New York, NY, USA, ACM Press (1987) 163–169

Phase-Based Registration of Multi-view Real-Time Three-Dimensional Echocardiographic Sequences

Vicente Grau[1], Harald Becher[2], and J. Alison Noble[1]

[1] Wolfson Medical Vision Laboratory, Department of Engineering Science, University of Oxford, Parks Road, OX1 3PJ Oxford, United Kingdom,
{vicente, noble}@robots.ox.ac.uk
http://www.robots.ox.ac.uk
[2] Department of Cardiology, John Radcliffe Hospital, Headley Way Headington, OX3 9DU Oxford, United Kingdom,
harald.becher@orh.nhs.uk

Abstract. Real time three-dimensional echocardiography opens the possibility of interactive, fast three-dimensional analysis of cardiac anatomy and function. However, at the present time these capabilities cannot be fully exploited due to the low image quality associated to this modality. We propose to increase image quality and information content by combining images acquired from different echocardiographic windows. In this paper, we present an algorithm to register these datasets. Phase-based measures have been proposed as a suitable alternative to intensity-based ones for ultrasound image analysis. The proposed algorithm uses a new cost function, based on local orientation and phase differences, to align the datasets. Visual observation of results, and preliminary numerical analysis, show the robustness and accuracy of this method.

1 Introduction

Among the imaging modalities that can be used for diagnosis of cardiovascular patients, echocardiography possesses a number of advantages that make it particularly useful: it is fast, low cost, portable and compatible with implanted cardiac devices and life-support equipment. However, the low quality of echocardiographic images when compared to cardiac MR is a serious drawback when calculating quantitative measures of anatomy and function. Furthermore, intensities depend on acquisition factors such as the incidence angle of the ultrasound beam and the depth within the tissue.

Recently, real-time three-dimensional echocardiography (RT3DE) has become available commercially. There is an agreement in the cardiology field that RT3DE is superior to two-dimensional echocardiography for the quantification of left ventricular volumes and function, but in spite of these advantages RT3DE has not yet become routine in the clinical practice due to difficult image acquisition, limited image quality and laborious manual data analysis (see e.g. [1]).

Echocardiographic acquisition can only be performed from certain probe locations or windows. The availability of three-dimensional echocardiographic data allows to

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 612 – 619, 2006.

combine datasets acquired from different windows. In [3], we showed the enhancement in image quality and information content that could be obtained by combining these datasets, assuming that the datasets were previously aligned. In Fig. 1, images acquired from the apical and parasternal windows, corresponding to the same heart plane, are shown. Big differences between the two images make the registration process extremely difficult if intensity-based approaches are used.

As mentioned above, the characteristics of ultrasound acquisition reduce the usefulness of intensity to identify different tissues. Previous research groups, including our own [2-4] have used phase-based algorithms with considerable success. The main advantage of phase-based algorithms is their invariance to image brightness and contrast, which makes them particularly suitable for ultrasound imaging.

In this paper, we demonstrate the feasibility of a phase-based algorithm to perform registration of RT3DE datasets acquired from the apical and parasternal windows. The user input required is limited to the selection of three landmarks on each of the datasets. The algorithm makes use of different image scales and is applied in a multiresolution framework to improve speed and robustness. Results obtained on 20 datasets from 3 subjects show a consistent, accurate behavior. Our contributions include, on the algorithmic side, the definition of a registration cost function based on local phase and orientation and calculated using the monogenic signal, and on the applications side, the proposal of a working compounding method to improve the information of RT3DE datasets by combining acquisitions from several windows.

2 Methods

2.1 Phase Calculation Using the Monogenic Signal

Several methods have been proposed for calculating phase in an N-dimensional signal. Among them, the monogenic signal [5], an isotropic extension of the analytic signal to N dimensions, provides a convenient and accurate calculation method. The monogenic signal of an N-dimensional signal is based on the Riesz transform, calculated by applying the set of N filters given, in the frequency domain, by

$$H_i(\boldsymbol{u}) = \frac{u_i}{|\boldsymbol{u}|}, \quad i = 1 \ldots N \tag{1}$$

where $u=[u_1 \ldots u_N]^T$, with u_i representing the ith coordinate unit vector (in 3D, $u_0=x$, $u_1=y$, $u_2=z$). To achieve frequency as well as spatial localization, a band-pass filter is applied to the signal prior to calculating the monogenic signal. In our case, we used log-Gabor filters.

The monogenic signal assigns an N+1-dimensional vector (the combination of the original signal value and the Riesz transform) to each spatial location. The N angles that define the vector can be used to estimate the N-D local signal orientation (θ) and phase (φ). Specifically, orientation and phase can be calculated as

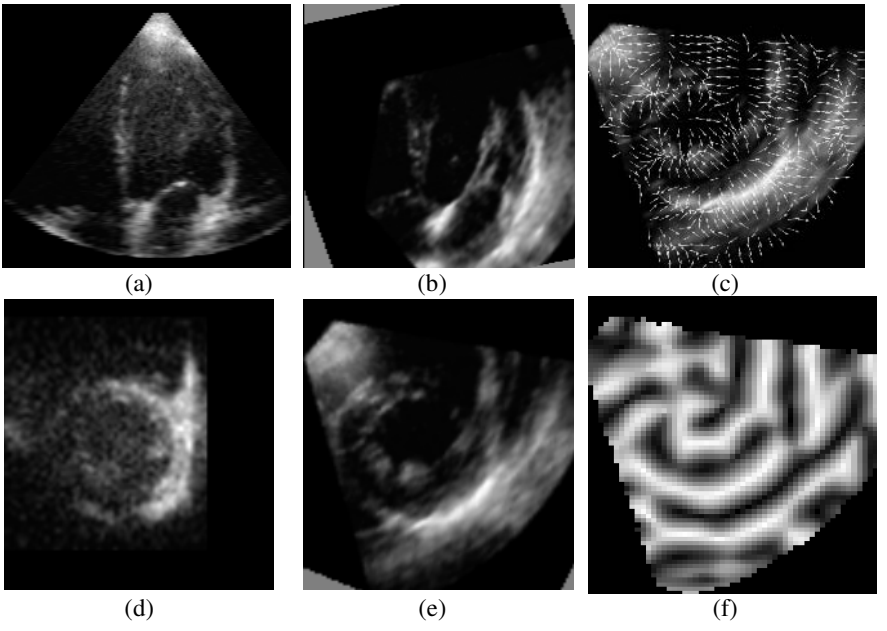


Fig. 1. Comparison of apical (a,d) and parasternal (b,e) datasets after registration. (a) and (b) correspond to long axis slices, while (d) and (e) are short axis. In (c) and (f), the orientation and phase of the short axis parasternal slice in (e) are shown. A mask corresponding to the ultrasound acquisition volume was used to avoid border effects.

$$o_i\left(I(x),s\right)=\frac{g_s(x)*h_i(x)*I(x)}{\sqrt{\sum_{i=1}^{N}\left(g_s(x)*h_i(x)*I(x)\right)^2}}\quad,\quad i=1\ldots N$$
$$\varphi\left(I(x),s\right)=\operatorname{atan2}\left(\sqrt{\sum_{i=1}^{N}\left(g_s(x)*h_i(x)*I(x)\right)^2},\quad g_s(x)*I(x)\right) \tag{2}$$

where o_i is the ith component of the normalized orientation vector, $g_s(x)$ is the band-pass filter with central wavelength s and $h_i(x)$ are the filters from (1), in the spatial domain. Confidence in the estimation of orientation angles depends on the value of the phase. We can quantify this confidence as $sin^2(\varphi)$ (more details are available in [5]).

Orientation and phase of the image in Fig. 1 (e), calculated after anisotropic filtering as explained in Section 2.3, are shown in the right column of Fig. 1. Even in low-contrast areas of the image, the algorithm returns a reasonable value for orientation. Notice that, at the central line of ridges, artefacts in orientation may appear. These are locations where the phase is close to zero, and thus have a low orientation confidence.

Our working hypothesis is that we can register multi-window RT3DE datasets by aligning local phase and orientation vectors calculated in (2). Sections 2.2 to 2.4 provide

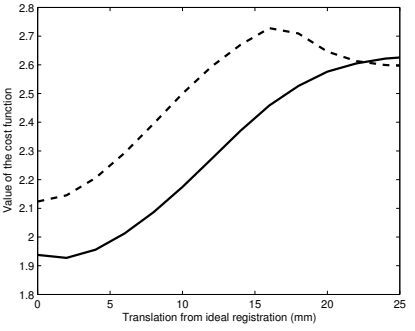


Fig. 2. Cost function values with respect to translation from the global minimum, for filters centered on wavelengths equal to 4 mm (dotted line) and 6 mm (solid line). Note that the image resolution is of the order of 1 mm. The capture range is wider in the case of the larger scale, while the lower scale achieves a higher accuracy. To obtain this figure, for each distance value d (horizontal axis), translations with norm *d* in different directions were calculated. The value shown in the figure corresponds to the minimum value of the cost function for each distance *d*.

more details about the complete registration algorithm. Section 3 shows the ability of the method to produce consistent, accurate results with limited user interaction.

2.2 Definition of a Registration Cost Function Based on Orientation Differences

The novel part of the applied algorithm is the definition of the cost function M(T). We propose to use an alignment of local orientation and phase, taking into account the confidence in orientation estimation, over the cardiac cycle:

$$M(T)=\sum_t \frac{\sum_x W(x,t)\cos\left(\left|\theta\left(I_1(x,t),s\right)-\theta\left(I_2\left(T(x,t)\right),s\right)\right|\right)}{\sum_x W(x,t)}+\cdots$$
$$\cdots+\sum_t \frac{\sum_x \varphi\left(I_1(x,t),s\right)-\varphi\left(I_2\left(T(x,t)\right),s\right)}{N_x} \quad (3)$$
$$W(x,t)=\min\left(\sin^2\left(\varphi\left(I_1(x,t),s\right)\right),\sin^2\left(\varphi\left(I_2(x,t),s\right)\right)\right)$$

with T being the geometrical transform mapping image I_2 to image I_1, φ and θ being phase and orientation, respectively, as introduced in Section 2.1, and $W(x,t)$ being the relative weights calculated from the orientation confidence. In (3), the cost function is estimated by accumulating the contributions of all spatial locations x and all frames t in the cardiac cycle. In general, calculation time can be reduced by using only a small set of frames. In our experiments, we used only the end-systolic and end-diastolic frames.

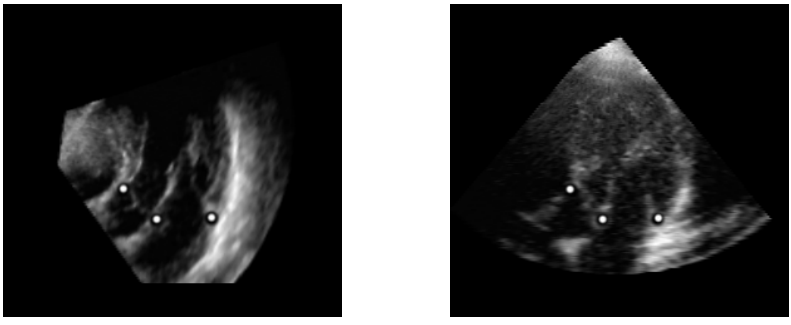


Fig. 3. Slices of the parasternal (left) and apical (right) datasets of a sample subject, with the points corresponding to the valve hinges highlighted

2.3 Registration Method

As we apply our method on multi-window datasets of the same patient, acquired with a time lapse of minutes, we do not expect a significant variation of shape and heart motion to appear between sequences. We thus use a rigid (translation + rotation) spatial transform. Regarding temporal registration, heart rate can change by a few frames between acquisitions, particularly when the subject moves to facilitate access to the scanning window. We perform this alignment by identifying the end-systolic and end-diastolic frames and performing piecewise linear temporal registration between them.

Minimization of the metric function (3) was done using Powell optimization, using only the end-systolic and end-diastolic frames in the calculation.

The influence of scale s is illustrated in Fig. 2: a larger scale enlarges the capture range of the metric and thus increases the probability of reaching the approximate location of the global minimum, while reduced scales provide a narrower minimum increasing the reproducibility of the final result when initialization landmarks vary. To combine these advantages, we embed the process in a multiscale scheme, by initially applying a large scale and using the result to initialize the process at the next, smaller scale. Furthermore, to increase robustness and decrease computation time, the whole process is carried out in a multiresolution framework. As a preprocessing step, spatial coherence-enhancing anisotropic diffusion [6] was applied.

The registration method was implemented using a combination of Matlab programs, to calculate phase and orientation, and ITK functions to perform the registration. New ITK classes were written to calculate the new cost function in (3).

2.4 Initialization

A reasonably good starting point is required to avoid the optimization method getting trapped in a local minimum. Though the cost function in (3) does not have a particularly narrow capture range (see Fig. 2), in the case of multi-window acquisition there is too much difference between the original datasets for the algorithm to converge.

Though one might calculate an approximate probe location, based on an estimation of approximate apical and parasternal positions for the general case, one would still have to estimate the approximate rotation of the probe about its axis, which in the 3D acquisitions can be modified at will by the clinician. We have thus chosen to use a landmark-based initialization. We identify three landmarks corresponding to the hinges of the mitral and aortic valves, selected at the plane crossing approximately through the centers of the mitral and aortic valve rings and parallel to the long axis of the left ventricle. In Fig. 3, the landmarks detected in a sample case are shown. Initialization is done by minimizing the sum of squared distances between landmarks.

In our experiments, the clinician could always locate the landmarks with enough accuracy. However, the complexity of the images makes this task difficult for a non-expert. Automatic initialization would be desirable and is a topic for further research.

3 Results and Conclusion

To evaluate the performance of the algorithm, 3 subjects were scanned at the John Radcliffe Hospital, Oxford, using a Philips Sonos 7500 scanner. Between 1 and 3 parasternal datasets and between 3 and 6 apical ones were obtained for each subject, changing the angle of the probe within each window to vary the information contained on each dataset. A total of 20 datasets were acquired in this way. For each dataset, the frames closest to end-systole and end-diastole were identified. Three landmarks were then selected for each one of the datasets, as described in Section 2.3.

Sample results for two of the subjects are shown in Fig. 4. It is clear that the algorithm succeeds in keeping continuity of the main structures. The full sequences are available at http://www.robots.ox.ac.uk/~mvl/htdocs/research/gallery.html.

Quantitative evaluation of image analysis techniques in RT3DE datasets is particularly complicated due to the difficulty in obtaining the ground truth. Using a 3D localizer during image acquisition is prone to errors caused by patient movement and displacement of the heart in the thoracic cavity due to patient breathing. Our own tests attaching the ultrasound probe to a mechanical arm (Faro Arm) showed alignment errors that were obviously larger than the ones in the results of the presented algorithm, thus making the procedure not suitable for ground truth estimation. Manual alignment of the datasets using 3D visualization software also proved impractical.

Manual segmentation of the datasets cannot be used to establish ground truth, as the quality of the original images is not sufficient to perform an exact manual delineation. As an alternative, an experienced cardiologist performed a visual assessment of the local accuracy of registration using the 16-segment model recommended by the American Society of Echocardiography to assess ventricular wall function. A diagram showing the names of the segments is shown in Fig. 5(a). Seven pairs of apical-parasternal datasets from 3 subjects, aligned using our method, were presented to the expert, who qualitatively scored the registration accuracy at each segment between 0 (worst) and 2 (best), with 2 meaning registration errors smaller than approximately 2 mm, 1 meaning errors between approximately 2 and 5 mm, and 0 bigger than approximately 5 mm (note that the resolution of the images is in the range of 1 mm).

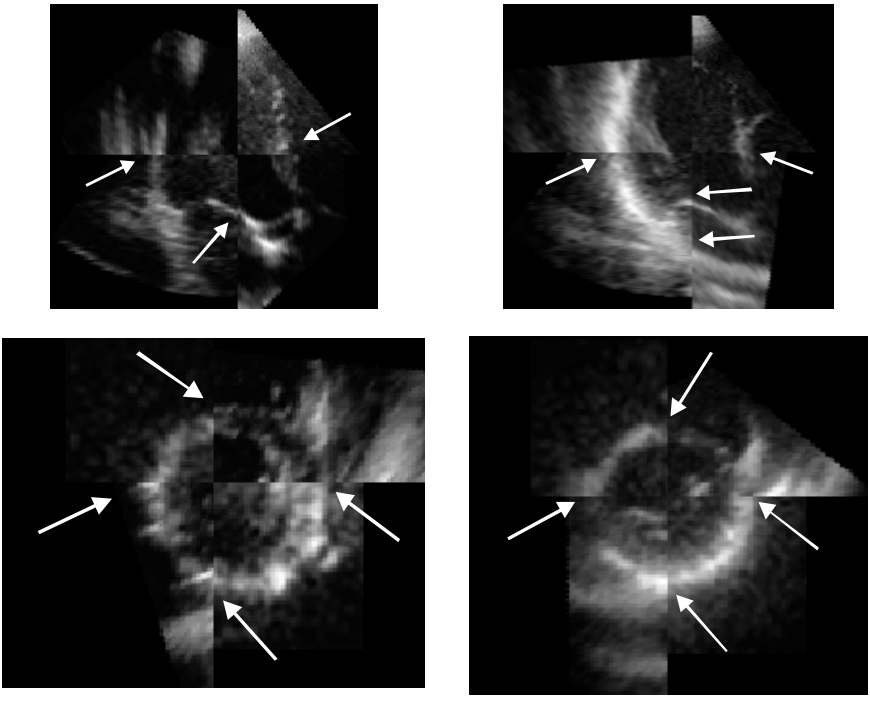


Fig. 4. Sample slices of two different subjects after alignment. Top and bottom rows show long and short axis cuts, respectively. Apical and parasternal datasets have been superimposed in a 2x2 checkerboard pattern. Notice continuity of the ventricular walls and valves at locations highlighted by the arrows.

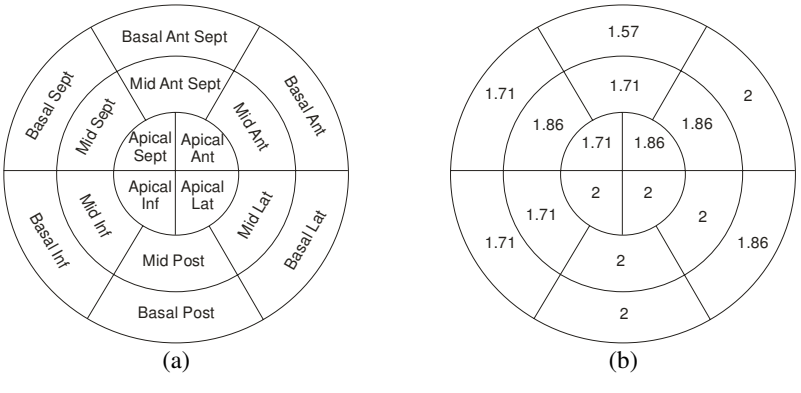


Fig. 5. Results of visual scoring of registration accuracy. (a) Names of the 16 segments in the model; (b) average score for each segment.

Average results for each segment in Fig. 5(b) show the accuracy of the method. Globally, 95 out of 112 individual segments received the maximum (2) score, 17 received a 1, and none received a 0 score. The global average score for all segments was 1.85.

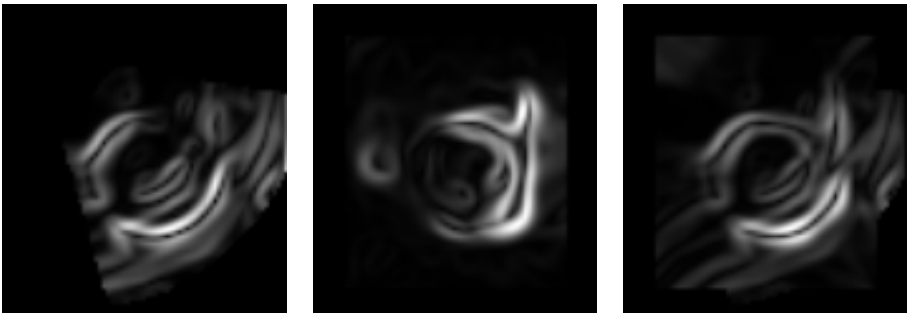


Fig. 6. Results of the Sobel filter on short axis slices from the parasternal (left) and apical windows (center), and on the combination of both (right). The completion of the circular pattern defining the left ventricle can be clearly appreciated.

The presented results, though obtained on a small sample, demonstrate the ability of the algorithm to produce accurate results for the apical-parasternal RT3DE registration problem. Fused images can be obtained using methods such as the one in [3]. In the end, the accuracy of our results will be validated by the improvement provided in real clinical tasks such as motion evaluation or volume measurements. In Fig. 6, we provide an illustration of the potential benefits of the newly generated images. The increased amount of edge information, corresponding to the endo- and epicardial surfaces, can be appreciated just by showing the result of a simple gradient estimation filter. We expect similar advantages for state of the art algorithms for motion estimation and ventricle segmentation.

Acknowledgment. The presented research was funded by the EPSRC grant GR/S72801.

References

1. Kühl, H.P. et al, "High-resolution transthoracic real-time three-dimensional echocardiography", Journal of the American College of Cardiology 43(11), 2083-2090, 2004.
2. Mulet-Parada, M., Noble, J.A., "2D+T Acoustic Boundary Detection in Echocardiography", Medical Image Analysis, 4(1):21-30 (2000)
3. Grau, V., Noble, J. A., "Adaptive Multiscale Ultrasound Compounding Using Phase Information", MICCAI 2005: 589-596
4. Hemmendorff, M. et al, "Phase-based multidimensional volume registration", IEEE TMI 21(12), 1536-1543, (2003)
5. Felsberg, Sommer, "The monogenic signal", IEEE TSP, 49(12):3136-3144, (2001)
6. Weickert, J., "Coherence-enhancing diffusion filtering," International Journal of Computer Vision, vol. 31, pp. 111–27, 1999.

Carotid Artery Segmentation Using an Outlier Immune 3D Active Shape Models Framework

Karim Lekadir and Guang-Zhong Yang

Visual Information Processing Group, Department of Computing
Imperial College London, United Kingdom
{lekadir, gzy}@doc.ic.ac.uk

Abstract. This paper presents an outlier immune 3D active shape models framework for robust volumetric segmentation of the carotid artery required for accurate plaque burden assessment. In the proposed technique, outlier handling is based on a shape metric that is invariant to scaling, rotation and translation by using the ratio of inter-landmark distances as a local shape dissimilarity measure. Tolerance intervals for each descriptor are calculated from the training samples and used to infer the validity of landmarks. The identified outliers are corrected prior to the model fitting using the ratios distributions and appearance information. To improve the feature point search, the method exploits the geometrical knowledge from the outlier analysis at the previous iteration to weight the gray level appearance based fitness measure. A combined intensity-phase feature point search is also introduced which significantly limits the presence of outliers and improves the overall search accuracy. Both numerical and *in vivo* assessments of the method involving volumetric segmentation of the carotid artery have shown that the outlier handling technique is capable of handling a significant presence of outliers independently of the amplitudes.

1 Introduction

Atherosclerosis is central to the clinical sequelae of coronary heart, cerebrovascular and peripheral vascular disease. For the assessment of atherosclerosis, cardiovascular MR is emerging as an attractive tool clinically because of its ability in identifying arterial remodeling, as well as providing information on plaque composition *in vivo*. The technique, however, is susceptible to motion and blood flow artifact and consistent plaque burden assessment involving the calculation of total vessel volume in response to therapeutic measures is a significant challenge. This is because small, progressive changes of the vessel volume over time require a high segmentation accuracy, and the traditional manual approach is subject to significant inter- and intra-operator variabilities. Most existing segmentation methods for the carotid artery segmentation [1-3] are semi-automatic and usually based on Active Contours. They require additional user interaction to correct for errors due to artifacts and the lack of a model to limit the segmentation results to valid instances.

Recently, automatic segmentation based on statistical shape modeling has attracted significant interests for vessel segmentation because of its unique strengths in exploit-

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 620 – 627, 2006.

ing *a priori* knowledge about the geometry of known anatomical shapes. The Active Shape Models (ASMs) [4], for example, can capture spatio-temporal principal shape variations from a set of labeled examples based on a Point Distribution Model (PDM). The model is then fitted to unseen shapes in an iterative manner where the pose and shape parameters are estimated using a least squares minimization approach. The ASM, however, is known to be problematic in the presence of image artifacts as well as incomplete or complex image features. In this case, some of the landmarks generated at the feature point detection stage are erroneous as they lie on incorrect boundary positions. These outlying landmarks are known to affect considerably the segmentation accuracy using the ASM search. A number of modifications have been introduced [5-8] to limit the effects of the outliers on the ASM search, but their actual performance is compromised as the number and amplitude of the outliers increase, especially for volumetric segmentation.

The purpose of this paper is to present a robust outlier handling method for volumetric segmentation with ASM [9]. The method detects and corrects the individual outliers prior to the estimation of the pose parameters by using an invariant shape metric based on the ratio of inter-landmark distances. Tolerance intervals are calculated at the training stage for each descriptor to detect invalid ratios, and thus the corresponding erroneous landmarks. Instead of rejecting the identified outliers or replacing them by the corresponding mean values, they are rectified by a combined use of the ratios distributions and appearance information. The geometrical knowledge gathered from the outlier analysis is propagated to successive iterations for feature point detection. This increases the search accuracy and reduces the number of iterations needed for final convergence. We also present a combined intensity-phase based feature point search, which significantly limits the presence of outliers and improves the overall accuracy. The practical value of the technique is validated by using a newly developed combined intensity-phase data acquisition scheme of plaque burden assessment for carotid arteries.

2 Methods

2.1 Outlier Handling

The ASM search involves in an iterative manner the detection of feature points [10] which are then fitted to the model by the estimation of pose and shape parameters. This is achieved using least squares minimizations which are known to be sensitive to extreme values. Thus, when outliers are present amongst the feature points, the ASM often introduce significant errors in the segmentation results. Without the correct estimation of the pose parameters, it is difficult to infer the presence of outliers due to residuals distribution introduced. Thus, the proposed outlier detection is based on a fully invariant shape descriptor using the ratio of interlandmark distances, which allows the decoupling of outlier analysis and ASM model fitting. This metric can be calculated for any triplet of points and represents the relative position of a landmark with regard to other points in the shape. An important property of this descriptor is its ability to perform as a geometrical measure of dissimilarity, which makes it ideal for

outlier detection. Each ratio r_{ijk} can be assumed to follow a Gaussian distribution and we use $\overline{r}_{ijk}$ and s_{ijk} to represent the mean and standard deviation of the distribution. The idea behind the proposed method is that outliers have some of their associated ratios as extreme values. Thus, outlier detection is achieved by using tolerance analysis [11], with which tolerance intervals are calculated for each ratio from the training samples by using the following equation:

$$T_{ijk} = \left[\overline{r}_{ijk} - k_2.s_{ijk}, \overline{r}_{ijk} + k_2.s_{ijk}\right] \tag{2}$$

where k_2 is the two-sided tolerance factor [9] representing the size of the interval. From each tolerance interval, a likelihood ratio measure f_r can be derived, *i.e.*,

$$f_r\left(r_{ijk}\right) = \begin{cases} 1 & \text{if } r_{ijk} \in T_{ijk} \\ 0 & \text{elsewhere} \end{cases} \tag{3}$$

This measure is equal to 0 if the ratio is an extreme value but it does not indicate which of the three landmarks associated with the ratio is invalid. By summing all the ratio likelihood measures that a given landmark is associated with, however, the likelihood of the point being an inlier can be estimated:

$$f_p\left(P_i\right) = \frac{1}{K_i}\sum_{j,k} f_r\left(r_{ijk}\right) f\left(r_{jki}\right) f\left(r_{kij}\right) \tag{4}$$

where K_i is the number of ratios associated with the landmark P_i.

Based on the above likelihood measure, an iterative algorithm is used where one outlier is identified at each iteration. When no outliers are present, all the likelihood measures of the feature points will be close to 1. In the presence of outliers, however, the likelihood measures decrease from 1, and the amount of decrease is more significant for the outliers. Thus, at each iteration, the landmark with the lowest likelihood measure is rejected and the likelihood measures of the remaining landmarks are updated by subtracting the contribution of the rejected point. This procedure is repeated until the lowest likelihood measure is close to 1, which suggests that all the remaining landmarks are inliers.

The derived inliers are then used to rectify the detected outliers by using a local search regime that maximizes the product of the p.d.f. of the associated ratios. This is equivalent to choosing the point that minimizes the following least-square function:

$$\sum_{j,k}\left(\frac{r_{ijk} - \overline{r}_{ijk}}{s_{ijk}}\right)^2 = \sum_{j,k}\left(\frac{d_{ij} - d_{jk}\overline{r}_{ijk}}{d_{jk}s_{ijk}}\right)^2 \tag{5}$$

where d corresponds to interlandmark distances. It should be noted that the correction suggested by Eq. (5) is only based on geometrical information and may not optimally correspond to local salient features. Therefore, a final local adjustment based on the gray level information is required. Subsequently, the residual distribution can be seen as Gaussian and the traditional ASM model fitting can be applied efficiently.

2.2 Improved Feature Point Detection

In this paper, two strategies are introduced to improve the feature point search. The first approach is directly connected to the outlier handling method described above and uses the geometrical knowledge gathered from the outlier analysis at previous iteration to weight the intensity fitness measure d_g as follows:

$$f_{total}\left(P_j\right) = d_g\left(P_j\right) \cdot w\left(P_j\right) \tag{6}$$

where the weights w take values from 0 to 1 and are calculated by using the point likelihood measure in Eq. (4) for each candidate point. This enables regions around the true position of the landmark point to be heavily weighted to avoid potential outliers from being generated. More importantly, this approach makes the segmentation procedure more stable as only a small number of iterations are needed for final convergence.

To further improve the system accuracy, feature point search based on combined intensity-phase information is also introduced. In cardiovascular MR, the assessment of plaque burden is usually achieved with 3D Turbo Spin-Echo (TSE) sequences. To suppress blood flow artifacts, double inversion preparation on either side of the excitation slab is introduced. Three-dimensional TSE techniques have the advantage of providing contiguous coverage and high SNR, but they can suffer from incomplete blood signal suppression in certain regions. However, it has been shown recently that by the introduction of velocity phase sensitivity to the sequence, it is possible to use the reconstructed phase images to separate blood signal from vessel wall boundaries [12]. This additional information can also be used for a more robust feature point search to further limit the presence of outliers. This is achieved by combining the gray level appearance with the phase information to construct a fitness measure for feature point search based on the phase profile φ:

$$\varphi = \left(\left|\varphi_1 - \varphi_s\right|, \ldots, \left|\varphi_h - \varphi_s\right|\right) \tag{7}$$

where h is the size of the profiles and φ_s the stationary phase (usually equal to 0). Both the intensity and phase profiles are then normalized as follows:

$$\hat{g}_i = \frac{g_i - g_{\min}}{g_{\max} - g_{\min}} \qquad \hat{\varphi}_i = \frac{\varphi_i - \varphi_{\min}}{\varphi_{\max} - \varphi_{\min}} \tag{8}$$

After normalization, the intensity and phase vectors are concatenated to form a combined profile vector p. PCA is then applied to the combined profile vectors of the training set for each individual landmark to extract the mean profile and the principal modes of variation. The Mahalanobis distance is then used as in the original ASM formulation [10] to calculate the global fitness measure as follows:

$$(p - \overline{p}) S_p^{-1} (p - \overline{p}) \tag{9}$$

Other strategies can also be used to combine appearance and phase information, such as the use of a *k*NN classification approach as proposed in [13].

2.3 Experiments

To assess the practical value of the algorithm, 25 carotid artery datasets were collected. Subjects with carotid artery atherosclerosis, as shown by prior carotid Doppler ultrasonography, underwent two CMR scans. Both asymptomatic and symptomatic subjects were included. MR was performed using a 1.5T scanner (Sonata, Siemens, Erlangen Germany), a purpose-built two-element phased-array surface carotid coil and a specially designed head and neck cushion for immobilization. A 3D volume-selective TSE sequence was used with a pixel size of $0.47mm^2$, slice thickness of 2 mm. A velocity phase sensitivity of $\pi/40$ rad/cm/s was used to acquire the phase images. For each dataset, 20 slices around the bifurcation were selected as a region for measurement. Delineations by an expert were performed using 272 landmarks. The data sets used for training and evaluation were selected on a leave-one-out basis. To tackle the problem of over-fitting and also to take into account the topological and geometrical configuration of the carotid artery including the main, internal and external branches, an ASM formulation similar to [14] was used such that points were grouped into segments at different cross sections of the branches.

3 Results

For a detailed assessment of the performance of the outlier handling algorithm in the presence of different levels of outliers, the following experiments were carried out by randomly adding synthetic outliers to the manual delineations using non-Gaussian noise. In a first experiment, the percentage of the outliers varied from 0 to 65% of the total number of landmarks while the amplitude was fixed to 50% of the average extent of the arteries. The average and standard deviation point to surface error in Fig. 1(a) demonstrate that the proposed method can handle up to 50 % of outliers. In a second experiment, the percentage of outliers was fixed to 25% while the maximum amplitude varied from 0 to 9.25mm. It can be seen from Fig. 1(b) that the performance of the method is independent of the extremity of the outliers.

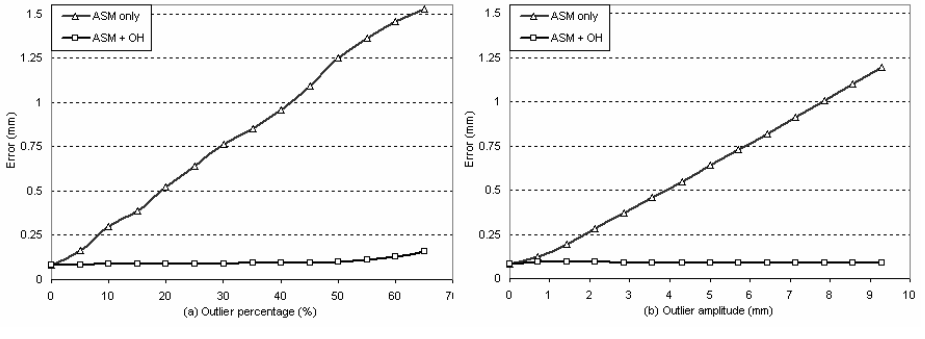


Fig. 1. Mean segmentation error of the traditional ASM and ASM with the proposed outlier handling (ASM+OH) in response to different percentage (a) and amplitude (b) of the outliers

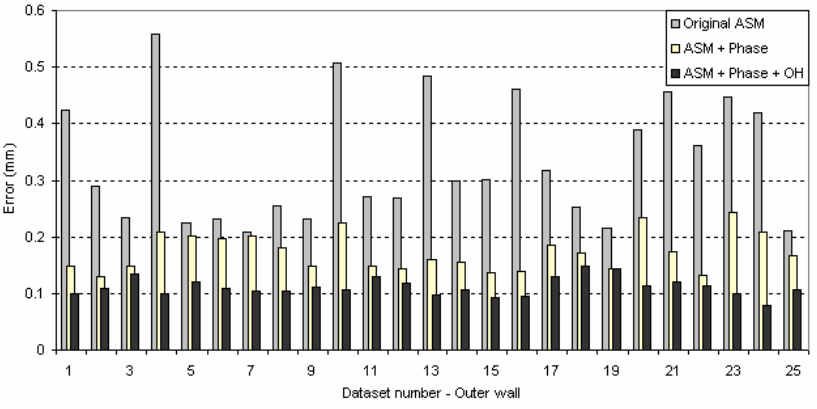


Fig. 2. Segmentation errors for the outer wall boundary by using the original ASM, ASM with the combined intensity-phase fitness measure (ASM+Phase), and the latter method with outlier handling (ASM+Phase+OH) for the 25 datasets used in this study

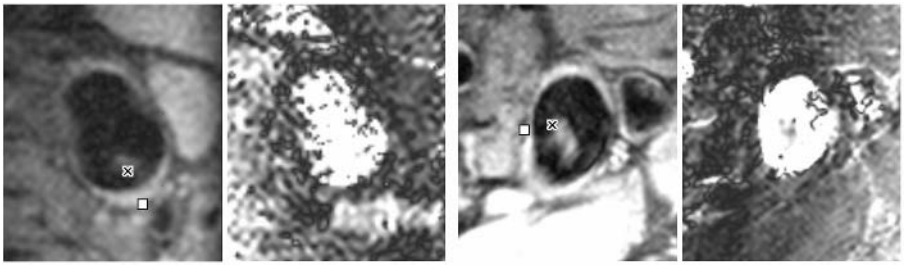


Fig. 3. 2D illustrations of the occurrence of the outliers (in crosses) due to blood flow artifacts present in the magnitude images and how the phase images at the right can be used for true boundary localization (in squares)

For *in vivo* validation, the original ASM and the proposed framework were applied to the 25 carotid artery datasets. Fig. 2 shows the segmentation results for the outer wall boundary which is particularly prone to outliers. Firstly, it can be seen that the combined intensity-phase feature point detection is considerably more robust as the average error decreases from 0.33 ± 0.12 mm for the original ASM to 0.17 ± 0.03 mm when using the proposed feature point search (47.91 % improvement). However, some outliers are still present amongst the feature points and further improvement is achieved using the proposed outlier handling scheme to a final average of 0.11 ± 0.02 mm. This corresponds to 63.28 % overall improvement of the original technique, which is significant for serial plaque burden assessment. Fig. 3 shows two examples taken at the bifurcation region, where outliers were generated by using the traditional intensity based fitness measure due to blood flow artifacts. It can be seen that the combined use of intensity and phase information is able to localize the correct boundary position. For 3D detailed visual illustration, Fig. 4 shows two examples of the application of the outlier handling procedure, followed by a robust generation of a

plausible instance of the shape by the ASM. The accuracy of the proposed method for the luminal and vessel outer wall boundary as well as the volumetric results are summarized in Table 1, demonstrating the achievable accuracy in a clinical setting.

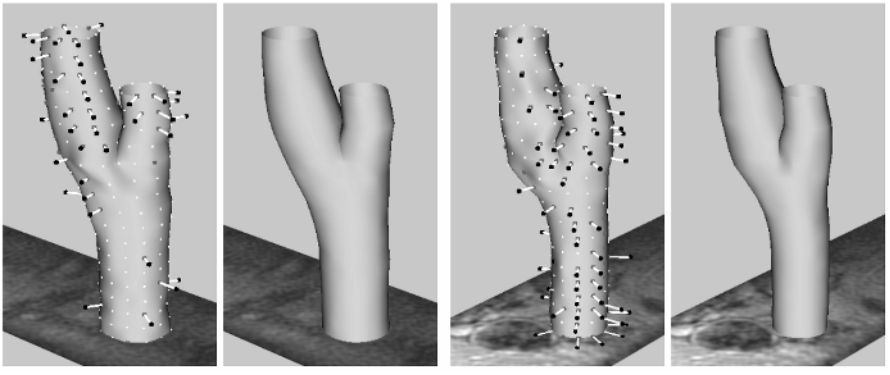


Fig. 4. Two 3D examples of the outlier handling algorithm, where the outliers (in black) are detected and corrected before the ASM model fitting procedure

Table 1. Overall segmentation accuracy of the patient data for the proposed technique

	Average Error	Standard deviation	Median Error	Maximum error
Luminal boundary (mm)	0.09	0.01	0.09	0.13
Outer wall boundary (mm)	0.11	0.01	0.10	0.14
Vessel wall volume (%)	3.19	2.43	2.13	8.64

4 Conclusions

In this paper, we have presented an outlier immune framework for volumetric segmentation that is suitable for plaque burden assessment. The method is based on outlier detection and correction prior to the model fitting procedure of the ASM using an invariant shape representation. The main advantage over exiting outlier handling techniques is that the proposed algorithm is independent of the severity of the outliers, and is robust even when half of the landmarks are incorrectly positioned. Furthermore, the technique uses a geometrically weighted fitness measure for feature point search, which exploits the result of outlier analysis from successive iterations, thus preventing outliers from re-appearing at subsequent iterations. Another contribution of the paper is the introduction of a combined intensity-phase scheme for feature point search, which limits the presence of outliers and reduces the average error of boundary localization significantly. The experimental results on CMR volumetric segmentation of the luminal and outer vessel boundaries of the carotid artery demonstrate the robustness of the algorithm and its potential clinical value.

Acknowledgements. The authors wish to thank N. Keenan and Prof. D. Pennell from the CMR Unit of Royal Brompton Hospital London for providing the carotid arteries data and the manual delineations.

References

1. Yang F, Holzapfel G, Schulze-Bauer C, Stollberger R, Thedens D, Bolinger L, Stolpen A, and Sonka M, "Segmentation of wall and plaque in in-vitro vascular MR images," *International Journal of Cardiovascular Imaging*, vol. 19, pp. 419-428, 2003.
2. Han C, Hatsukami TS, Hwang JN, and Yuan C, "A fast path minimal path active contour model," *IEEE Transactions on Medical Imaging*, vol. 10, pp. 865-873, 2001.
3. Ladak HM, Thomas JB, Mitchell JR, Rutt BK, and Steinman DA, "A semi-automatic technique for measurement of arterial wall from black blood MRI," *Medical Physics*, vol. 28, pp. 1098-1107, 2001.
4. Cootes TF, Cooper D, Taylor CJ, and Graham J, "Active shape models - Their training and application," *Computer Vision and Image Understanding (CVIU)*, vol. 61, pp. 38-59, 1995.
5. Rogers M and Graham J, "Robust active shape model search," presented at European Conf. on Computer Vision (ECCV), 2002.
6. Duta N and Sonka M, "Segmentation and interpretation of MR brain images: An improved active shape model," *IEEE Transaction on Medical Imaging*, vol. 17, pp. 1049-1067, 1998.
7. Behiels G, Vandermeulen D, Maes F, Suetens P, and Dewaele P, "Active shape model-based segmentation of digital X-ray images," presented at Proc. Medical Image Computing and Computer Assisted Intervention (MICCAI), 1999.
8. Li H and Chutatape O, "Boundary detection of optic disk by a modified ASM method," *Pattern Recognition*, vol. 36, pp. 2093-2104, 2003.
9. "Outlier detection and handling for robust 3D active shape models search," *IEEE Transaction on Medical Imaging*, submitted for review.
10. Cootes TF and Taylor CJ, "Active shape model search using local grey-level models: A quantitative evaluation," presented at Proc. British Machine Vision Conf. (BMVC), 1993.
11. Guttman I, *Statistical tolerance regions: Classical and Bayesian*. London: Griffin, 1970.
12. Crowe LA, Varghese A, Mohiaddin RH, Yang G-Z, and Firmin DN, "Elimination of residual blood flow-related signal in 3D volume-selective TSE arterial wall imaging using velocity-sensitive phase reconstruction," *Journal of Magnetic Resonance Imaging*, vol. 23, 2006.
13. Ginneken Bv, Frangi AF, Staal JJ, Romeny BMtH, and Viergever MA, "Active shape model segmentation with optimal features," *IEEE Transaction on Medical Imaging*, vol. 21, pp. 924-933, Aug. 2002.
14. Davatzikos C, Tao X, and Shen D, "Hierarchical active shape models, using the wavelet transform," *IEEE Transaction on Medical Imaging*, vol. 22, pp. 414-423, Mar. 2003.

Estimation of Cardiac Hyperelastic Material Properties from MRI Tissue Tagging and Diffusion Tensor Imaging

Kevin F. Augenstein[1], Brett R. Cowan[2],
Ian J. LeGrice[1], and Alistair A. Young[1,2]

[1] Bioengineering Institute, University of Auckland, New Zealand
[2] Center for Advanced MRI, University of Auckland, New Zealand
a.young@auckland.ac.nz

Abstract. The passive material properties of myocardium are important in the understanding of diastolic cardiac dysfunction. We determined hyperelastic myocardial material parameters in four isolated arrested pig hearts undergoing passive inflation of the left ventricle. Using geometry from MRI, recorded boundary conditions, muscle fiber architecture from diffusion tensor imaging, and deformation from tissue tagging, finite element models were constructed to solve the finite elasticity stress estimation problem. The constitutive parameters of a hyperelastic transversely isotropic material law were determined by minimizing the difference between the predicted and imaged deformation field. The optimized parameters were in a similar range as those reported by previous studies, showing increased passive stiffness in the muscle fiber direction. The average RMS error was 0.92 mm, similar to the image resolution of 0.80 mm. Optimization of hyperelastic models of myocardial mechanics can thus be performed to extract meaningful biophysical parameters from MRI data.

1 Introduction

A better understanding of the motion and mechanical function of the heart in normal and disease states would result in more accurate and effective assessment and diagnoses of cardiac disease in a clinical environment. Current technologies provide a great deal of information about heart function, as there are a large range of imaging modalities (including echocardiography, CT, and MRI) available. However, a lack of applicable and simple models prevents this data from being fully utilized. This is due, in part, to the complex relationship between cardiovascular disease processes and the wide range of observed shapes and motions achieved by the heart. Typically, the data do not densely cover the full region of interest (e.g., MRI provides high resolution data within the imaging plane but relatively low temporal and through-plane resolution). The application of mechanical modeling methods allows interpolation of the sparse data sets over the domain, and characterization of the underlying biophysical phenomena.

In particular, estimation of the compliance of heart tissue is important for the understanding of diastolic dysfunction [1]. Stiffening of the myocardium leads

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 628–635, 2006.

to impairment of filling which in turn leads to decreased pump function via the Frank-Starling mechanism. Previous methods proposed to estimate material properties from image data have required limiting assumptions on the form of the material law (often linear elasticity is assumed) or the equations of motion (often small strain theory is imposed) [2, 3, 4]. These assumptions are known to be violated in the heart, which undergoes large strains (20% shortening and up to 100% wall thickening) and exhibits nonlinear elastic behavior [5].

We investigated methods for the determination of hyperelastic material parameters from MRI data. In order to fully characterize the boundary conditions applied to the heart, an isolated arrested porcine heart testing protocol was employed. MRI was used to determine the geometry of the heart, the tissue deformation from SPAMM tissue tagging, and the muscle fiber architecture from diffusion tensor imaging. The left ventricle (LV) was cyclically inflated using a computer controlled stepper motor driven pump, with simultaneous LV pressure measurement. The finite elasticity problem was solved using the finite element method, resulting in estimation of stress given a set of material parameters, and the known boundary conditions. The finite elasticity problem was embedded into a parameter optimization loop, which determined the best match between imaged and modeled deformations. This protocol has been previously validated using a silicon gel phantom [6]. Results are presented for four hearts.

2 Method

2.1 Experimental Protocol

Four pigs were initially anesthetized with zoletil, and maintained with 2.5% halothane in oxygen. The right femoral artery was cannulated to monitor arterial pressure and a pigtail catheter was inserted into the left ventricle via the right carotid artery. A median and left lateral thoracotomy was performed to expose the heart. The left main coronary artery and right coronary artery were dissected near their origins, and ties were loosely placed for coronary occlusion after arrest. The animal was heparinized (100 IU kg^{-1}) via the right carotid artery catheter. A snare was placed around the aorta and the heart was arrested by injection of cold potassium citrate (60 ml, 15%) into the left ventricle via the catheter. Simultaneously, the aortic snare was tightened, thereby perfusing the heart with potassium citrate via the coronary arteries and causing immediate arrest in a relaxed state. The heart was then rapidly removed and rinsed in cold saline, before being perfused with 2,3-butanedione monoaxime (200 ml, 50 mM in saline) through cannulae inserted into the right and left coronary ostia. The coronary arteries were then occluded using the previously placed snares to stop any subsequent flow through the aorta entering the coronary network.

A slit was made in the left atrium to allow access to the mitral valve. A large-bore cannula with an annular disc on the end was introduced through the mitral valve, and a thread looped around the cordae tendineae, and tied tight to the cannula above the disc. A second disk was then screwed down on to the valve leaflets clamping them between the upper and lower discs and forming a seal.

The aortic valve was destroyed and a large diameter cannula introduced in the aorta and sealed with a snare about the aorta. The heart was then suspended in a Plexiglas box (160 mm cube) filled with physiological saline solution, which was clamped into a head coil for imaging. The mitral cannula was attached to a computer controlled reciprocating pump and the aortic cannula attached to the pressure line. Repeatable inflation cycles of approximately 16 ml and a period of 1.2 s were performed during imaging and pressure measurement [6].

2.2 Imaging

SPAMM tagged images were acquired (field of view 200 mm, 6 mm slice thickness, 5 deg flip angle, 128 x 256 matrix, 16 phases, and 5 mm stripe spacing in a grid pattern) at 6 short axis positions, and at 6 long axis slices at 30 degree increments about the central axis of the left-ventricle. A segmented k-space imaging sequence was used (seven lines per segment, TR 188 ms, TE 4 ms, 2 averages), requiring 38 deformation cycles per image slice to acquire.

In order to obtain an independent measure of the ventricular inflow and out flow during the deformation cycle, phase contrast velocity magnetic resonance images were acquired at the level of the mitral inlet (field of view 150 mm x 150 mm, 5 mm slice thickness, 256x256 matrix, 64 phases), prospectively gated on every alternate R-wave so that the entire cycle was acquired. These allowed calculation of volume inflow and outflow through the mitral cannula throughout the deformation cycle.

Diffusion tensor imaging was performed with the heart at zero transmural pressure, in order to obtain information on the tissue micro-structure. A pulsed gradient spin echo diffusion sequence with echo planar imaging and slice readout (two-dimensional multi-slice acquisition, 12 contiguous axial slices, 5 mm slice thickness, TR 14589.0 ms, TE 100.06 ms, no averages, in plane spatial resolution of 0.82 mm, and slice thickness of 5 mm) was used to acquire diffusion-weighted images as described in [7]. Six diffusion gradient direction images (along with a zero weighted image) were acquired per slice, with diffusion weighting b values of 500, 1000, 1500 and 2000 s mm-2 in each direction, resulting in a data set of 28 images per slice.

2.3 DTI Analysis

For each pixel in the region of interest the six independent components of the diffusion tensor D_{ij} and the non-diffusion weighted intensity ($A(0)$) were estimated using Levenberg-Marquardt nonlinear regression. For each pixel, the following objective function was minimized:

$$\min_{A(0),D_{1\ldots3,1\ldots3}} \sum_{i=1}^{m}\sum_{j=1}^{n} \frac{1}{\sigma_j} \parallel A(0)e^{-b^{ij}:D} - A(b^{ij}) \parallel_2^2 \qquad (1)$$
$$st\, D_{ij} = D_{ji}\ i,j = 1\ldots3$$

where m is the number of diffusion directions (7), n is the number of gradient magnitudes (4), b^{ij} is the b-matrix with a given diffusion direction and weighting, and $A(b^{ij})$ is the image intensity of the pixel. Weights for the 28 residuals (4 magnet strengths times 7 diffusion directions) were set to be $1/\sigma_i$, where σ_i was the standard deviation of the image intensity in free water for the given diffusion gradient magnitude. The dominant diffusion direction was given by the eigenvector associated with the largest eigenvalue of the diffusion tensor. In the heart, this corresponds to the mean muscle fiber direction in the imaged voxel [8].

2.4 Finite Element Modeling

To represent the geometry and the muscle architecture, a finite element model was constructed in 3D rectangular Cartesian coordinates. The geometry and fiber angle field was interpolated using tri-cubic Hermite basis functions, with the element coordinates chosen to lie in the circumferential, longitudinal, and transmural directions respectively. The boundaries of the heart were determined from the untagged images using guide point modeling [9]. Tags were tracked throughout the cycle and 3D displacement fields were reconstructed for each frame using previously described methods [10]. The fiber field nodal parameters were optimized to fit the largest eigenvector of the DTI diffusivity tensor.

A transversely-isotropic Fung-type exponential constitutive law was used to characterize the material behavior, as described previously [11, 12, 13, 14]:

$$W = \frac{1}{2}C\left(e^{b_1 e_{ff}^2 + b_2(e_{rr}^2 + e_{ss}^2 + 2e_{sr}e_{rs}) + 2b_3(e_{fs}e_{sf} + e_{fr}e_{rf}) - 1}\right) \tag{2}$$

C, b_1, b_2, and b_3 are material constants and e_{ij} are components of the Lagrangian strain tensor, e, referred to the local fibre (f), transmural or radial (r) and cross-fibre (s) coordinates. This material law implies material symmetry about the muscle fibre direction, f. Optimization of the material parameters of this law has been performed in several species using various experimental protocols (Tab. 1).

The motions of the basal nodes were constrained to be the same as the imaged motion. All other nodal displacements were unconstrained in the stress solution process. The pressure trace was averaged over every cycle in the imaging sequence, and then pressures for individual image frames were calculated by binning all recorded pressures over the imaging time.

Table 1. Material parameters for the transmural Fung exponential hyperelastic strain energy function, obtained by previous studies

Paper	Species	C(kPa)	b_1	b_2	b_3
Guccione and McCulloch [11]	Canine	0.88	18.5	3.58	1.63
Okamoto *et al.* [14]	Canine	0.3913	75.41	33.86	11.50
		1.0173	77.49	8.016	3.122
		0.7476	39.53	6.174	5.665
		0.3047	49.89	11.11	14.63
		0.1008	93.03	61.61	73.10

The material parameters were optimized by solving the finite elasticity equations with a given set of material parameters and boundary conditions. The predicted motion of the tag points was then compared with their imaged motion to calculate an error for each frame:

$$e = \sum_{i=1}^{N} \parallel x_i - x(\boldsymbol{\xi}_i, \mathbf{C}) \parallel_2^2 \tag{3}$$

where N is the number of data points, x_i is the reconstructed deformed data point, and $x(\boldsymbol{\xi}_i, \mathbf{C})$ is the geometric interpolation function for the deformed geometry, at a given material point, x, and for a given set of material parameters, $\mathbf{C}$.

Sequential Quadratic Programming (SQP) [15] was used to minimize the error with respect to the material parameters. The first derivatives, with respect to the material parameters, were calculated using a finite difference scheme. This involved solving the finite elasticity problem with each perturbed material parameter. In this case, the stress solution was restarted from the unperturbed stress solution in order to reduce solution calculation times. A combined relative and absolute tolerance of 0.001 of the L2 norm of the change in the material parameters was used as the convergence criterion for the optimization. The Hessian and covariance matrix of the objective function was calculated at the optimum solution, in order to derive information about the correlation between the material parameters.

2.5 Validation

The method and apparatus have been previously validated [6]. Briefly, a silicone gel (Sylgard 511 dielectric gel) phantom was constructed in the shape of a cylindrical annulus. The silicon gel was well described by a neo-Hookian material law with a single material parameter $C1 = 8.71 \pm 0.06kPa$ estimated independently using a rotational shear apparatus. The MRI derived C_1 parameter was allowed to vary regionally and was estimated as $8.80 \pm 0.86kPa$ across the model. Monte Carlo simulations showed that acceptable errors in material parameters (less than 5%) were obtained using five frames of the inflation cycle, and noise levels on the order of a pixel (0.8 mm standard deviation).

3 Results

The geometry and deformation fields were reconstructed in the finite element model to an RMS error of approximately the pixel size (0.8 mm). Angular errors in the fiber field ranged from 8 to 13 degrees, comparable to 17 degrees in previous histological studies [16]. Figure 1 shows the transmural variation of the fitted fiber field in the model. We found that a finite element mesh with eight elements circumferentially, eight elements longitudinally and two elements transmurally gave converged stress solutions and good numerical convergence of the finite elasticity equations.

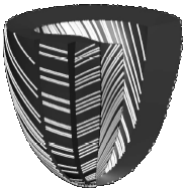


Fig. 1. Transmural fiber field fitted to DTI data at varying depths from epicardial to endocardial surfaces

Table 2. LV pressures at five frames in the inflation cycle for the four experiments

Pressures (kPa)
0.44 0.84 1.24 1.49 1.81
0.54 0.78 1.07 1.40 1.51
0.32 0.65 0.81 1.32 1.60
0.39 0.89 1.31 1.54 2.02

Table 2 shows pressures from five frames of the inflation cycle in the four experiments. These frames were input into the parameter optimization process. Previous validation experiments showed that five frames were sufficient to characterize the material parameters to within 5% error [6]. Table 3 shows the fitted material parameters for the four hearts used in this study, along with the total error, number of data points and the RMS error. The RMS error is approximately the same size as the pixel size in the original tagged images. A sample correlation matrix is given in Table 4. The condition number for the covariance matrix for the four experiments ranged from 2723 to 6078. Sample short-axis and long-axis images are shown for each of the studies in Fig 2. Superimposed on the images are the model displacements (as crosses). The arrows show the tracked tag points from the reference frame (least pressure) to the frame of highest pressure.

4 Discussion

Finite element models of geometry and micro-structure were constructed from MRI derived data. These models, together with tracked material points and

Table 3. Optimal material parameters and RMS errors for the four experiments

C	b_1	b_2	b_3	error	No. data points	RMS error (mm)
0.7945	8.03	6.12	8.21	3076.43	2450	1.1206
0.2238	20.65	6.18	20.65	3310.23	3765	0.9377
0.0695	83.43	36.37	62.36	1709.12	3005	0.7542
0.3112	27.18	10.19	23.82	4116.07	5410	0.8723

Table 4. Typical correlation matrix at the optimal solution

	Correlation			
	C	b_1	b_2	b_3
C	1.0000	0.2062	-0.6408	-0.1129
b_1		1.0000	0.0313	0.0618
b_2			1.0000	-0.5429
b_3				1.0000

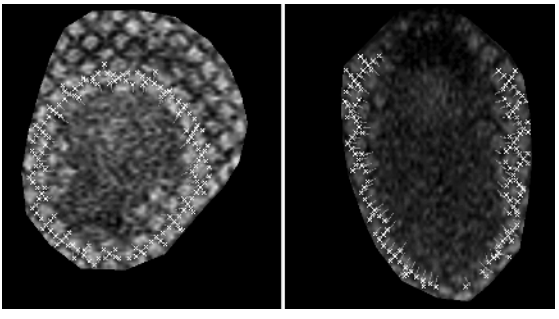


Fig. 2. Short and long axis model displacements (arrows) and predicted tag stripe locations (white points)

recorded pressures, were used to estimate material parameters. One novel aspect of this experimental protocol was that the papillary muscles were left attached to the mitral valve leaflets which were tied. However, in the model, the papillary muscles were excluded and their effects on the mechanical response is yet to be determined. The fitted parameters differed between experiments, but fell within the range previously obtained in the literature (Table 1).

The large condition number of the Hessian indicates that parameters of this law are not well defined by the experimental data, although the fitted parameters were capable of reconstructing the tracked material points with an RMS of 0.92 mm, approximately equal to the image resolution of 0.8 mm. The choice of the appropriate material law formulation for heart myocardium, including allowing for possible regional variations of parameters, is currently an open research topic. [17] have proposed alternative formulations based on strain attributes which are chosen to minimize the correlation between terms in the material law. Myocardial material laws based on this formulation show promise for material parameter estimation from magnetic resonance displacement data, by reducing the correlation between fitted parameters (as found in the current study) and thereby improving the accuracy of the parameter estimation.

In order to extend this work to characterize the material properties of the heart in-vivo, the left ventricle filling pressure and other boundary conditions must be non-invasively determined. Computational fluid dynamics may be used to determine noninvasive pressure gradients in late filling, which may then be combined with strain measurements [3].

References

1. Mandinov, L., Eberli, F.R., Seiler, C., Hess, O.M.: Diastolic heart failure. *Cardiovasc. Res.* **45** (2000) 813–825
2. Hu, Z., Metaxas, D., Axel, L.: In vivo strain and stress estimation of the heart left and right ventricles from mri images. *Medical Image Analysis* **7** (2003) 435–444
3. Wen, H., Bennett, E., Epstein, N., Plehn, J.: Magnetic resonance imaging assessment of myocardial elastic modulus and viscosity using displacement imaging and phase-contrast velocity mapping. *Magn Reson Med* **54** (2005) 538–548
4. Sermesant, M., Rhode, K., Sanchez-Ortiz, G., Camara, O., Andriantsimiavona, R., Hegde, S., Rueckert, D., Lambiase, P., Bucknall, C., Rosenthal, E., Delingette, H., Ayache, N., Razavi, R.: Simulation of caridac pathologies using an electromechanical biuventricular model and xmr interventional imaging. *Medical Image Analysis* **9** (2005) 467–480
5. Yin, F.: Ventricular wall stress. *Circulation Research* **49** (1981) 829–842
6. Augenstein, K., Cowan, B., LeGrice, I., Nielsen, P., Young, A.: Method and apparatus for soft tissue material parameter estimation using magnetic resonance imaging. *J Biomech Eng.* **127** (2005) 148–157
7. Basser, P., Pierpaoli, C.: A simplified method to measure the diffusion tensor from seven mr images. *Magn Reson Med* **39** (1998) 928–934
8. Garrido, L., Wedeen, V., Kwong, K., Spencer, U., Kantor, H.: Anisotropy of water diffusion in the myocardium of the rat. *Circ Res.* **74** (1994) 789–93
9. Young, A., Cowan, B., Thrupp, S., Hedley, W., Dell'Italia, L.: Left ventricular mass and volume: fast calculation with guide-point modeling on MR images. *Radiology* **216** (2000) 597–602
10. Young, A., Kraitchman, D., Dougherty, L., Axel, L.: Tracking and finite element analysis of stripe deformation in magnetic resonance tagging. *IEEE Trans Medical Imaging* **14** (1995) 413–421
11. Guccione, J.M., McCulloch, A.D., Waldman, L.K.: Passive material properties of intact ventricular myocardium determined from a cylindrical model. *ASME Journal of Biomechanical Engineering* **113** (1991) 42–55
12. Guccione, J.M., McCulloch, A.D.: Finite element modeling of ventricular mechanics. In Glass, L., Hunter, P., McCulloch, A.D., eds.: Theory of Heart - Biomechanics, Biophysics, and Nonlinear Dynamics of Cardiac Function, Springer-Verlag, New York (1991) 121–44
13. Emery, J.L., Omens, J.H., McCulloch, A.D.: Strain softening in rat left ventricular myocardium. *ASME Journal of Biomechanical Engineering* **119** (1997) 6–12
14. Okamoto, R.J., Moulton, M.J., Peterson, S.J., Li, D., Pasque, M.K., Guccione, J.M.: Epicardial suction: A new approach to mechanical testing of the passive ventricular wall. *Journal of Biomechanical Engineering* **1122** (2000) 479–487
15. Gill, P.E., Murray, W., Wright, M.H. In: Practical Optimization, (Academic Press, New York)
16. P. M. F. Nielsen, I. J. Grice, B.H.S., Hunter, P.J.: Mathematical model of the geometry and fibrous structure of the heart. *American Journal of Physiology* **260** (1991) H1365–H1378
17. Criscione, J.C., McCulloch, A.D., Hunter, W.C.: Constitutive framework optimized for myocardium and other high-strain, laminar materials with one fiber family. *Journal of the Mechanics and Physics of Solids* **50** (2002) 1681–702

Boosting and Nonparametric Based Tracking of Tagged MRI Cardiac Boundaries

Zhen Qian[1], Dimitris N. Metaxas[1], and Leon Axel[2]

[1] Center for Computational Biomedicine Imaging and Modeling, Rutgers University, New Brunswick, New Jersey, USA
[2] Department of Radiology, New York University, New York, New York, USA

Abstract. In this paper we present an accurate cardiac boundary tracking method for 2D tagged MRI time sequences. This method naturally integrates the motion and the static local appearance features and generates accurate boundary criteria via a boosting approach. We extend the conventional Adaboost classifier into a posterior probability form, which can be embedded in a particle filtering-based shape tracking framework. To make the tracking process more robust and faster, we use a PCA subspace shape representation to constrain the shape variation and lower the dimensionality. We also learn two shape-dynamic models for systole and diastole separately, to predict the shape evolution. Our tracking method incorporates the static appearance, the motion appearance, the shape constraints, and the dynamic prediction in a unified way. The proposed method has been implemented on 50 tagged MRI sequences. The experimental results show the accuracy and robustness of our approach.

1 Introduction

Tagged cardiac magnetic resonance imaging(MRI) is a well known technique for non-invasively imaging the detailed myocardial motion and deformation throughout the heart cycle. In the tagged MRI protocol, ECG triggering is used to acquire a time sequence of images at several frames in a heart cycle. Quantitative analysis of the myocardial wall motion requires accurate segmentation of the epicardial and the endocardial boundaries for the whole time sequence. This segmentation problem is difficult due to the presence of tagging lines, the complexities of the cardiac geometry, the fast motion of the myocardium, and the high level of MR image noise. There have been some prior efforts to achieve tagged MRI segmentation. In [1], grayscale morphological operations were used to separate non-tagged blood-filled regions from the tagged myocardium regions. In [2, 3], learning methods including shape, intensity and appearance modeling were proposed.

It has been noted by several researchers that incorporating temporal information would greatly help the segmentation. When a human expert manually delineates the contours, it is common practice to watch the MRI images as a *movie* and use other frames and their motion to segment the current frame. Using this observation, in this paper we propose a dynamic segmentation approach. Several methods in the MRI literature have used temporal information to help segmentation. Sun *et al.* [4] proposed a segmentation and tracking

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 636–644, 2006.

method for the left ventricle by learning the ventricle dynamics. In [5], tracking of the myocardium is embedded in a nonrigid image registration framework. More myocardial boundary tracking methods can be found in the echocardiography literature. Comaniciu *et al.* [6, 7] proposed a multi-model information fusion framework to achieve robust myocardial boundary tracking. In [8, 9] shape-space based or contour-matching based approaches are proposed.

In this paper, we introduce a new framework for boundary tracking of short axis (SA) tagged MRI sequences, including the boundary contours of the endocardium of the left ventricle (LV), the endocardium of the right ventricle (RV) and the epicardium. In this framework, in order to find strong image features, instead of visually tracking the contour points via profile matching or optical flow related methods, we learn a more complex boundary criterion using AdaBoost. Contour profile consistency between two consecutive frames (the motion feature) is an important feature in tracking, but it has limitations in handling the inherent appearance changes due to tag fading and myocardial deformation, especially the boundary's tangential movement (such as rotation and sliding, see Fig. 2b) and myocardial motion through the imaging plane. Using the boosting algorithm, we demonstrate that for more accurate boundary feature extraction, the local boundary appearances in the current frame (the static feature) are also important. The AdaBoost algorithm gives a natural way to combine both features and generate better boundary criteria. Furthermore, through boosting error analysis, we find the confidence ratings of each criterion and calculate the posterior probability density function of the shape model. In order to lower the dimensionality and constrain the shape variations, we project the shape into a Principal Component Analysis (PCA) subspace [10], and update the PCA parameters instead of the contour points' positions. Since the heart motion has a cyclic contraction and expansion pattern, which exists in both normal and diseased hearts, we treat the systole and diastole separately and find two motion prediction matrices and error covariance matrices via multivariate linear regression. We embed the shape tracking into a nonparametric-based sequential particle filtering framework [11] for its ability of contour tracking through heavy clutter.

Our overall learning and tracking framework is outlined by the flow-chart in Fig. 1. Our experimental results show the accuracy and effectiveness of the proposed algorithm.

2 Shape Modeling and Boosting the Boundary Criteria

2.1 Shape Model

The shape of the mid portion of the heart in SA images is consistent and topologically fixed (one LV and one RV). Therefore it is reasonable to implement a PCA subspace shape model [10] to represent prior shape knowledge and constrain the shape variations during the myocardial wall tracking process. This is particularly desirable in the tagged cardiac MRI case. For many reasons including the noisy nature of tagged MRI and the complex structure of the heart, *accurate* boundary tracking based solely on the MR images is usually not possible. See, for example,

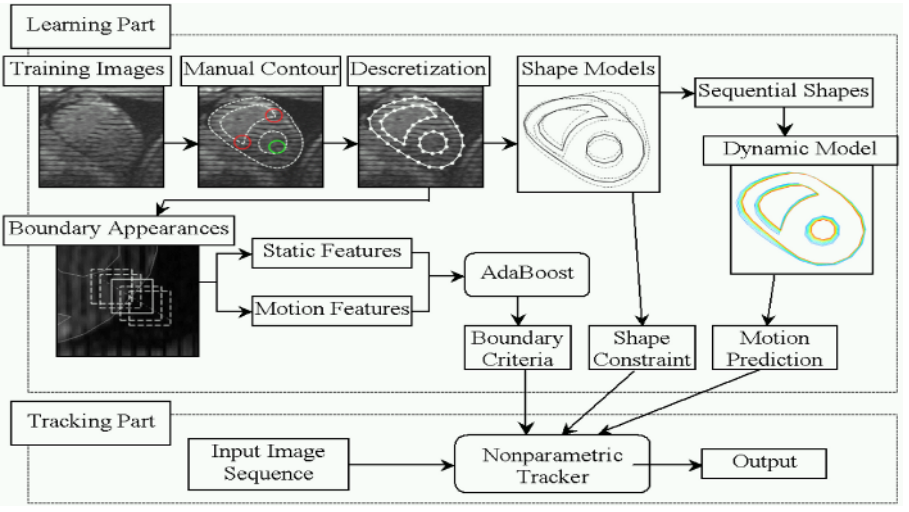


Fig. 1. The flow chart of our proposed framework illustrates the learning and tracking processes. In the manual contour block, the red circles indicate the cusp points between the septum and the RV endocardium. The green circle indicates the location of papillary muscle. In the dynamic block, cooler color turning into warmer color indicates the dynamic model's evolution during systole.

the manual contour box in Fig. 1, which shows an expert's manual segmentation results: the two cusp points between the septum and the RV endocardium are always kept, and the papillary muscles in LV and RV are excluded from the endocardium. This cannot be done automatically without strong shape priors.

We use the expert's manual contours as the PCA input. The manual contours are first centered and transformed to the same scale and orientation. Then they are automatically discretized into 50 ordered landmark points (see the discretization box in Fig. 1) using geometry features such as maximum curvature. Then principal component analysis is applied and the modes of shape variation are found. Any heart shape can be approximately modeled by $X = \bar{X} + Pb$, where $\bar{X}$ is the mean shape vector, P is the matrix of shape modes, and b is the vector of shape parameters weighting the shape variations. The following boundary criterion learning method will be applied on the landmark points individually.

2.2 Learn Boundary Criteria Using Adaboost

Probability Estimation in Adaboost. To reliably and accurately track the myocardial boundary, we try to use all possible image cues, including the static features and the motion features, to make a strong boundary criterion. This cue-integration is performed by Adaboost. Adaboost is a well known classification algorithm [12] that incrementally selects a small number of important features from a huge potential feature set and creates a linear combination of them as an accurate strong classifier. In the learning stage, given training data

$(x_1, y_1)...(x_m, y_m)$, where $x_i \in X$, $y_i \in \{0, +1\}$, during the boosting process, each iteration selects one feature, i.e., a weak classifier, h_t from the total potential features pool, and combines it (with an appropriate weight α_t) with the existing classifier that was obtained in the previous iterations.

$$h_t : X \rightarrow \{0, +1\} \tag{1}$$

$$F = \Sigma_t \alpha_t h_t(x) \tag{2}$$

After t iterations, the linear combination of the selected important features F makes a stronger classifier $H(x)$ with higher accuracy. We denote $\Gamma = \Sigma_t \alpha_t / 2$ as the threshold value for F. For a new input data x to be classified, if $F > \Gamma$, H classifies x as positive, and vice versa.

$$H(x) = \frac{Sign(F - \Sigma_t \alpha_t / 2)}{2} + \frac{1}{2} \tag{3}$$

We perform Adaboost boundary learning at each landmark point on the shape model. Thus the boundary criteria vary according to each model point's location. In order to embed the Adaboost $\{0, 1\}$ classifier into the sampling-based tracking framework, we need to extend it to a probability representation. Influenced by [13], we use a logistic function to estimate the probability of a point x_k, along with its local appearance patch, being on the boundary at the location of landmark point k, with a normalization term Γ_k the same as the threshold value:

$$Pr(y_k = +1|x_k) = \frac{e^{F_k(x_k)/\Gamma_k}}{e^{F_k(x_k)/\Gamma_k} + e^{-F_k(x_k)/\Gamma_k}} \tag{4}$$

After the training process, Adaboost provides the error rates e_k on the testing data. We define the confidence rating of the boundary criterion at point k as:

$$c_k = \ln \frac{1}{e_k}; \tag{5}$$

Intuitively, when c_k is big, we trust its probability prediction $Pr(y_k = +1|x_k)$ more. We incorporate the confidence rating into the probability $Pr(y_k = +1|x_k)$ as a posteriori knowledge. Then the probability density function of the shape model is estimated by:

$$p(Y = +1|X) \propto \prod_k \frac{e^{F_k(x_k)c_k/\Gamma_k}}{e^{F_k(x_k)c_k/\Gamma_k} + e^{-F_k(x_k)c_k/\Gamma_k}} \tag{6}$$

Boundary Feature Design. The boosting algorithm investigates a large number of weak feature candidates. At each landmark point on the shape contour, the local appearance features are captured by filtering methods in both static and motion cases.

To capture the static local appearance features, we design three different kinds of steerable filters. We use the derivatives of a 2D Gaussian to capture the edges, we use the second order derivatives of a 2D Gaussian to capture the ridges, and we use half-reversed 2D Gabor filters to capture the tagging line breakpoints.

The half-reversed 2D Gabor filters are defined as a 2D sine wave multiplied with the one directional derivative of a 2D Gaussian:

$$F = G'(x, y) \cdot \mathbb{R}\{e^{-j[\phi+2\pi(Ux+Vy)]}\} \tag{7}$$

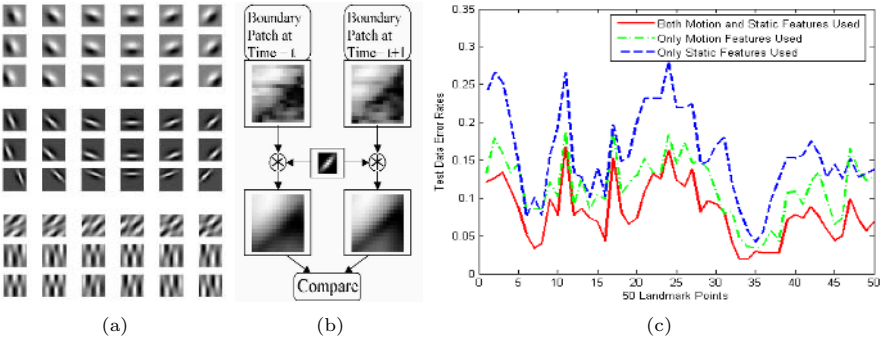
where G' is the derivative of a 2D Gaussian. U and V are the frequencies of the 2D sine wave, and ϕ is the phase shift.

To capture the motion local appearance features, we measure the intensity consistency between two consecutive frames. In order to avoid the effect of boundary tangential motion, before comparison, a set of Gaussian filters are designed to blur the boundary local patch in a certain orientation and scale (See Fig. 2(b).

For a 15x15 sized filter patch, by tuning the filters' scale, orientation and frequency, we designed 1840 static filters and 121 motion filters in total. See Figure 2(a) for some sample static filters.

In the manually segmented images, at each landmark point of the contours, a small image patch around it is cut out as a positive appearance training sample for this particular landmark point. Then, along the normal of the contour, on each side of the point, we cut out another two patches as the negative appearance training samples for this particular landmark point. (See boundary appearance box in Fig. 1) For motion features, the consistencies are measured between the current frame sample with the positive sample in the previous frame.

After the training process, the classification error rates on the testing data give the measurement of the boosting performance. After 50 Adaboost iterations, we find that the error rates using both static and motion features are much lower than using either static or motion features alone. See Figure 2(c). This comparison validates our boundary tracking scheme based on both motion and static features.



(a) (b) (c)

Fig. 2. (a) shows 3 sample sets of static feature filters, from top to bottom, the derivatives of Gaussian used for edge detection, the second derivatives of Gaussian used for ridge detection, and the half-reversed Gabor filters used for tag line breakpoint detection. (b) illustrates the function of the motion feature filter, which helps remove the tangential motion. (c) shows the testing error rates of the boundary criteria boosting. The red curve which integrates both motion and static features has lower error rates than the other two.

3 Nonparametric Shape Tracking

Sampling-based tracking algorithms have the ability to handle contour tracking problems with non-Gaussian probability density functions (pdfs). Our tracking method is based on a nonparametric particle filtering approach [11], where the pdf is represented by factored sampled particles.

During myocardial wall tracking, we use a state vector s to represent the shape model. $s = [c, \theta, t_x, t_y, b]$, where c is a scaling factor, θ is the orientation angle, (t_x, t_y) is translation, and b is a 12-D variation vector of the PCA shape modes. The myocardial wall tracking algorithm based on particle filtering is as follows:

1. Suppose the total sample number is N, total time frame number is T. At $time = 1$, $s_1^{(n)}$ is initialized as the manual contour shape at the first frame. Weights $\pi_i^{(n)} = \frac{1}{N}$.
2. For $time = 2 : T$,
   Factored draw samples $s'^{(n)}_t$ from s_{t-1} based on the weights $\pi_{t-1}^{(n)}$.
   Predict new $s_t^{(n)}$ by sampling from a dynamic model $p(x_t|x_{t-1}) = s'^{(n)}_t$.
   Measure and weight the new shape in term of the measurement y_t by Eq. 6. $\pi_t^{(n)} \propto p(Y_t|X_t = s_t^{(n)})$, and $\sum_n \pi_t^{(n)} = 1$
   Estimate the solution via the strongest local mode.
   End for

There are two main difficulties in implementing the algorithm. The first one is how to find a proper dynamic model. The second is how to decide the sampling size N.

The motion pattern differs greatly among human hearts, especially between the normal and the diseased. Our strategy is to find a common dynamic model whose constraint is loose but can help in decreasing the sampling range. Since the tagged MRI sequences are acquired with an ECG trigger, the end systolic (ES) frame is easy to locate. We train the dynamic model based on systole and diastole separately. For a given state vector $s^{(t)} = [s_1^{(t)}, s_2^{(t)}, ..., s_n^{(t)}]'$ at time t, we make an assumption that it can be approximated by a linear transform of $s^{(t-1)}$.

$$s_i^{(t)} = a_0 + a_1 s_1^{(t-1)} + a_2 s_2^{(t-1)} + ... + a_n s_n^{(t-1)} + N(0, r_i^2) \tag{8}$$

This is a multivariate linear regression problem. We solve it to get $A_i = [a_0, a_1, ..., a_n]$ as the prediction vector of s_i and r_i^2 as the error variance. Then the conditional probability of the prediction dynamic model is:

$$p(s_i^{(t)}|s^{(t-1)}) = \frac{1}{\sqrt{2\pi} r_i} \exp\{-(s_i^{(t)} - A_i \cdot s^{(t-1)})^2 / 2r_i^2\} \tag{9}$$

The sampling size N is experimentally set. For 12 sequences of the testing data, we find the tracking typically stops improving after $N \geq 1000$. Thus we set $N = 1000$.

4 Experimental Settings and Results

We acquired 50 time sequences of short-axis tagged MR images from 5 normal subjects and 3 patients. Each sequence contains 12 to 18 frames (images). In total, we collected 776 images. The spatial positions of these SA images vary from near the ventricle apex to near the ventricle base, but their topologies, only one LV and one RV, are consistent. The tagging line orientations are either 0°, 90° or ±45°. An expert was asked to segment the epicardium, the LV endocardium and the RV endocardium from the images. The manual segmentations were used as the training data, as well as the ground truth in validation based on a leave-6-out scheme, in which each time 44 sequences were grouped as the training data while the other 6 were used for testing. If two sequences were taken from the same subject and at the same spatial position, but with different tag orientations, we grouped them in the same training or testing set. In this way, we strictly separated the training and testing data.

Each training image is rotated and scaled to contain a 80x80-pixel image of the heart with the interior chest wall on top. The size of the boundary appearance patches is set to 15x15 pixels. In the tracking process, the initial contour is set semi-automatically [3] allowing manual correction. Our algorithm implementation is coded in Matlab 6.5 and runs on a 3GHz PC workstation. For a tracker with a sampling size of 1000, it takes 60 seconds to track a frame on average.

See Fig. 3 for some representative tracking results. Since the resulting contours are represented by 50 landmark points, we measure the tracking error by calculating the distances from the landmark points s_i to the expert's manual

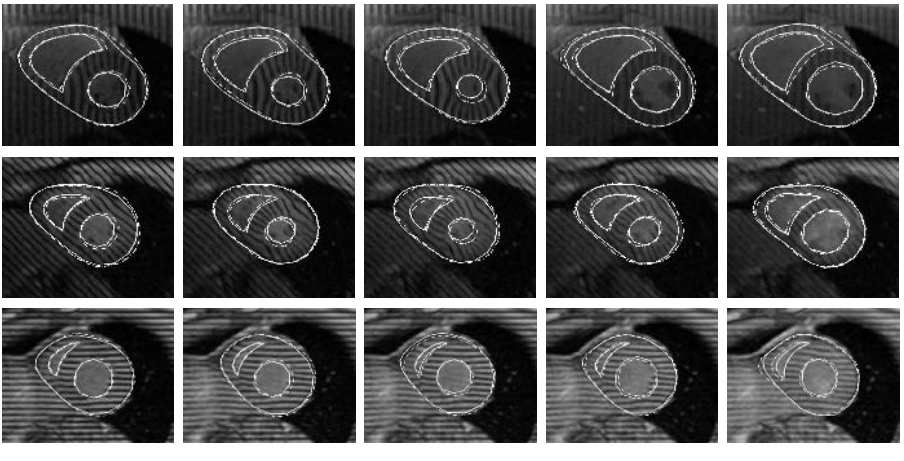
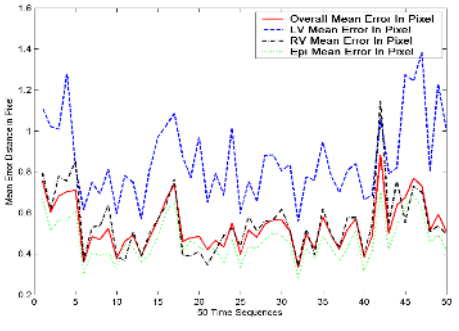


Fig. 3. Snapshots from three example sequences of our experimental results. The solid contours are from our tracking method, while the dashed are from manual segmentations. The first 2 rows are data from 2 normal subjects. The last row is from a patient with heart failure. Note that our tracking method can exclude the papillary muscle from the ventricle endocardium, and always keep the cusp points at the RV-LV junctions.

contour C: $error(s_i, C) = min_{c_i \in C} \parallel s_i - c_i \parallel$. Fig. 4(a) shows the mean error in pixels for each time sequence; the mean error for the LV. RV and epicardium are also drawn. In Fig. 4(b) some error analysis are performed in millimeters.

Conclusion/Discussion. In this paper, we have proposed a tracking framework for tagged MRI. Our method integrates the boundary appearance (both static and motion), the shape constraints and the dynamic model naturally in a boosting and nonparametric tracking framework. We strictly test the algorithm on data that are excluded in the training set. The experimental results show the accuracy. With the error analysis shown in Fig. 4(b), we find that in general we have achieved sub-millimeter accuracy while MRI accuracy is approximately $1mm$. Our method works slightly better on the mid-ventricle slices and on normal hearts. In the future we will train more shape and motion models for different slice positions and extend up to the valve region, where the RV separates into distinct inflow and outflow regions. We will also extend the method by training on different diseases, where the shape and motion of the patient data can be very different from the normal (see the last row of Fig. 3, which is from a patient with heart failure).



(a)

	$\overline{error}$ (mm)	$\overline{\sigma}$ (mm)
LV	1.24	0.66
RV	0.79	0.81
Epi	0.67	0.91
Near Base	0.77	0.89
Mid	0.71	0.81
Near Apex	0.87	0.99
Normal	0.76	0.86
Patient	0.89	1.03

(b)

Fig. 4. (a) The mean error for each time sequence measured in pixel. The red solid curve represents the overall error. The blue dashed, black dash-dot and green dotted curves represent the mean error distances in the LV, RV endocardium and epicardium respectively. (b) Error analysis in millimeters.

References

1. Montillo, A., Metaxas, D., Axel, L.: Automated segmentation of the left and right ventricles in 4d cardiac SPAMM images. In: MICCAI. (2002) 620–633
2. Huang, X., Li, Z., Metaxas, D.N.: Learning coupled prior shape and appearance models for segmentation. In: MICCAI (1). (2004) 60–69
3. Qian, Z., Metaxas, D., Axel, L.: A learning framework for the automatic and accurate segmentation of cardiac tagged MRI images. In: Proceedings Of CVIBA Workshop, In Conjuction with ICCV, LNCS 3765. (2005) 93–102

4. Sun, W., Cetin, M., Chan, R., Reddy, V., Holmvang, G., Chandar, V., Willsky, A.: Segmenting and tracking the left ventricle by learning the dynamics in cardiac images,. In: Proc. of Information Processing in Medical Imaging. (2005) 553–565
5. Chandrashekara, R., Mohiaddin, R., Rueckert, D.: Analysis of 3-d myocardial motion in tagged MR images using nonrigid image registration. IEEE Trans. Med. Imaging **23**(10) (2004) 1245–1250
6. Comaniciu, D., Zhou, X.S., Krishnan, S.: Robust real-time myocardial border tracking for echocardiography: an information fusion approach. IEEE Trans. Med. Imaging **23**(7) (2004) 849–860
7. Georgescu, B., Zhou, X.S., Comaniciu, D., Rao, B.: Real-time multi-model tracking of myocardium in echocardiography using robust information fusion. In: MICCAI (2). (2004) 777–785
8. Jacob, G., Noble, J., Behrenbruch, C., Kelion, A., Banning, A.: A shape-space-based approach to tracking myocardial borders and quantifying regional left-ventricular function applied in echocardiography. IEEE Trans. on Medical Imaging **21**(3) (2002) 226–238
9. II, J.C.M., Duncan, J.S.: Shape-based tracking of left ventricular wall motion. IEEE Trans. Med. Imaging **16**(3) (1997) 270–283
10. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active shape models - their training and application. Computer Vision and Image Understanding **61**(1) (1995) 38–59
11. Isard, M., Blake, A.: Contour tracking by stochastic propagation of conditional density. In: ECCV (1). (1996) 343–356
12. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. In: EuroCOLT '95: Proceedings of the Second European Conference on Computational Learning Theory. (1995) 23–37
13. R. E. Schapire, Y. Freund, P.B., Lee, W.S.: Boosting the margin: a new explanation for the effectiveness of voting methods. Annals of Statistics **26**(5) (1998) 1651–1686

A Region Based Algorithm for Vessel Detection in Retinal Images

Ke Huang[1] and Michelle Yan[2]

[1] Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI48824, USA*
[2] Siemens Corporate Research, 755 College Road East, Princeton, NJ08540, USA

Abstract. Accurate retinal blood vessel detection offers a great opportunity to predict and detect the stages of various ocular and systemic diseases, such as glaucoma, hypertension and congestive heart failure, since the change in width of blood vessels in retina has been reported as an independent and significant prospective risk factor for such diseases. In large-population studies of disease control and prevention, there exists an overwhelming need for an automatic tool that can reliably and accurately identify and measure retinal vessel diameters. To address requirements in this clinical setting, a vessel detection algorithm is proposed to quantitatively measure the salient properties of retinal vessel and combine the measurements by Bayesian decision to generate a confidence value for each detected vessel segment. The salient properties of vessels provide an alternative approach for retinal vessel detection at a level higher than detection at the pixel level. Experiments show superior detection performance than currently published results using a publicly available data set. More importantly, the proposed algorithm provides the confidence measurement that can be used as an objective criterion to select reliable vessel segments for diameter measurement.

1 Introduction

The retinal vessels are the only part of the central circulation that can be optically observed *in vivo*. It offers an opportunity to non-invasively explore the relationship of systemic microvascular disease to cardiovascular disorders. For example, signs of retinopathy, e.g. microaneurysms, retinal hemorrhages and soft exudates, are the symptoms of microvascular damage from elevated blood pressure, diabetes, and other processes [1]. Changes in width of retinal vessels are the indication of both ocular diseases and cardiovascular disorders, such as glaucoma [2], hypertension [3], diabetic retinopathy [4] and congestive heart failure [5]. Retinal vessel diameters are often measured in a concentric zone surrounding the optic disc, yielding the value of arterio-venous ratio (AVR), an independent and significant risk factor measure used for examining many systemic diseases [6]. Lower AVR is commonly interpreted as an indicator of

* This work was completed when K. Huang was an intern at Siemens Corporate Research.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 645–653, 2006.

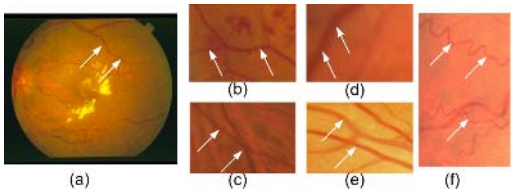


Fig. 1. Samples of an original retinal image (a) and different appearances of vessels. The large variation in appearance is caused by: (b) hemorrhage; (c) black background texture; (d) motion blurring; (e) different depths, diameters, contrast and central flex in the middle of arteries; (f) tortuosity. Arrows in the figure annotate the locations of some vessels.

arteriolar narrowing that is strongly associated with the onsets and effects of systemic diseases. In addition to diameter measurement, vessel detection is also used for retinal image registration [7], the optic disk localization [8], and as prior structural information in lesion detection [9].

Vessel detection in retinal images is a very challenging task due to several reasons. In the imaging process, retinal vessel structures at different depths are projected and overlapped onto the two-dimensional retinal image. A white strip called central reflex often stands out in the center of an artery and misleads the detection process. Uneven illumination, blurring, tortuosity and lesions of various appearances may also be present in the images. The width of vessel varies in a large range, from several pixels to tens of pixels. Fig. 1 illustrates an original retinal image with bright lesions and different appearances of retinal vessels. Physiological structures with blood are best contrasted in the green channel of color retinal images [10, 11, 12, 13]. Therefore, green channel image is used for vessel detection in the paper.

1.1 Related Work

Existing approaches for retinal vessel detection generally fall into two categories: tracking and classification. Thresholding can be viewed as classification with a simple classifier. An example of classification is introduced in [12], where the responses of an image to a set of matched filters are analyzed with threshold probing. In [13], ridge lines are first identified and more than 20 features are defined and computed for each pixel. A general process for tracking approaches starts with identifying some seed points, and the vessel structures recursively grow to their neighboring pixels from the seed points. In [14], a tracking algorithm following this paradigm is presented, where matched filters are used to analyze the neighboring pixels of currently detected vessels for recursive tracking. In [11], a tracking algorithm following similar steps is also proposed and verified. An algorithm combining thresholding and classification is proposed in [15], where the gray values of pixels are thresholded with decreasing values and results are refined by a classification process with a feature vector measuring properties such as angle, width, contrast and size. A combination of tracking

and classification is introduced in [10], where vessel detection is implemented by combining the tracking algorithm introduced in [14] and classification with a 6-dimensional feature vector that represents the likelihood of a pixel being inside a vessel.

1.2 Our Proposed Method

The blood vessel often appears to be a dark and continuous linear structure in retinal images. Normally, the boundaries of a short vessel segment are parallel to each other. To capture these visually salient properties, a region based Bayesian detection method is proposed for vessel detection in this paper, which works as follows. Cross sections of a vessel are first identified by searching boundary points along two parallel edges of an edge map. A confidence value is computed for each cross section to reflect its contrast to the background. The identified cross sections are then grouped into potential segments along the direction of a vessel. Each segment is associated with a confidence value based on the segment length and the variance of cross sections in width. Bayesian rule is introduced to combine the confidence values obtained at each step to generate a final confidence map, where a higher value represents a higher confidence of the segment being part of a vessel. By exploring the salient properties of retinal vessel at higher than the pixel level, contextual information can be effectively incorporated into the detection process. The confidence map in the proposed method offers a potential reference in the application where it needs to know how reliable of each detected vessel segment is. To the best of our knowledge, no current retinal vessel detection provides such a measurement.

The rest of this paper is organized as following. Section 2 discusses the implementation of region based Bayesian detection. Experimental results are presented in Section 3 and Section 4 concludes this paper.

2 Region Based Bayesian Detection

The proposed region based Bayesian detection is guided by first obtaining the quantitative measurements of the shape and continuity property, and then combining the measurements in the framework of Bayesian decision. Retinal vessels often appear to be darker in gray value than their neighbor regions, which implies that the intensity is a local property to be measured. Therefore, applying a single threshold to gray values of an entire image can not reliably separate vessel and non-vessel regions. To address this problem, shade correction [9] is used to compensate for the background variations across the whole retinal image. A conservative threshold is applied to the image obtained from the shade correction to generate a binary mask B, where the pixels with the value higher than the threshold are excluded from the following processing, as shown in Fig. 2.

2.1 Shape Measurement

An edge map is obtained by applying Canny edge detector to the green channel image. Since boundaries of a short vessel segment are normally parallel and close

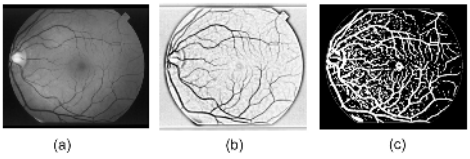


Fig. 2. Generating binary mask with shade correction: (a) green channel image; (b) image after shade correction; (c) binary mask B, where only the white regions are searched for potential vessels.

to each other, edges that are parallel to each other are first identified in the edge map. To achieve this goal, the edge map is searched horizontally and vertically. Taking searching row r horizontally as an example: denote P_i and P_j as two neighbor edge points on row r, and P_i and P_j are considered as a pair of vessel boundary points only when both of the following conditions are satisfied:

1. $(P_j - P_i) > D_l$ or $(P_j - P_i) < D_h$
2. $\mu(r, P_i, P_j) < \mu(r, P_i - D_t, P_i - 1)$ and $\mu(r, P_i, P_j) < \mu(r, P_j + 1, P_j + D_t)$

The first item introduces the constrain on vessel width. The retinal image used in the experiment is with a size of 605×700. D_h is set to 25 pixels and D_l is set to 2 pixels to avoid missing any potential vessels. In the second item, the function $\mu(r, P_i, P_j)$ denotes the mean value of image intensity of the pixels between P_i and P_j in row r of the corresponding green channel image. This constraint requires that the appearance of vessel pixels should be less or darker than its neighbor regions. D_t is defined as the number of the neighbor pixels at each side and is set to 5 in the experiment. Once P_i and P_j are identified as a pair of vessel boundary points, a value $C1$ is empirically defined to reflect the degree of the confidence associated with the cross section between P_i and P_j using the following formula:

$$C1 = (E_i + E_j)\, e^{-\alpha_1 |E_i - E_j|} \tag{1}$$

where E_i and E_j are the edge intensity at the point P_i and P_j, respectively. A larger E_i or E_j reflects a stronger image contrast between the vessel and its neighbors. However, because the two boundary points at a cross section should be close to each other in distance, a large difference in intensity between one boundary point to the other should be penalized. In the experiment, α_1 is set to 0.05 for this penalization. All pixels along the cross section between P_i and P_j are assigned to the value $C1$. The pixels not on any identified cross section are set to 0. In this way, a 2D confidence map S_h is obtained for shape measurement in the horizontal direction. The confidence map S_v for vertical shape measurement can also be obtained in a similar way by searching the images vertically. Fig. 3(a) and (b) demonstrate the shape confidence measurement in the horizontal and vertical directions, respectively.

The gradient information that is available in the edge detection process may be used for searching boundary points at cross sections. In the ideal case, the searching direction should follow the local gradient of each vessel boundary. How-

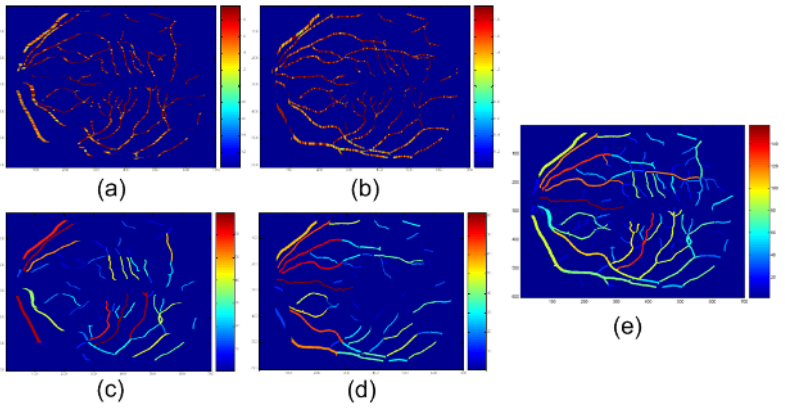


Fig. 3. Measured confidence maps of shape and continuity: (a) shape confidence map S_h in the horizontal direction; (b) shape confidence map S_v in the vertical direction; (c) continuity confidence map C_h in the horizontal direction; (d) continuity confidence map C_v in the vertical direction; (e) the final confidence map obtained using Bayesian rule. Red color represents high values and blue color represents low values.

ever, due to the noise and discrete representation of an image, not all boundary points can be uniquely paired and the searching lines may not cover all pixels inside each vessel segment. Therefore, we choose to horizontally and vertically search the image to make sure that every pixel is searched and a vessel segment at any orientation can be detected.

2.2 Continuity Measurement

In addition to the vessel structures, other non-vessel structures may also generate strong response in either or both of S_h and S_v. One such example is the background region between two neighboring bright lesions, as shown in Fig. 4. The salient property that could potentially separate the vessels from the rest is that the vessel often has an extended linear and continuous structure. This observation motivates us to further explore the contextual information by quantitatively measuring the continuity of the vessel.

The continuity is measured based on the cross sections that are identified in the section 2.1. The basic operation is to recursively connect each single cross section to form a continuous vessel segment, i.e., a vessel segment is detected in a similar fashion to the region growing, but using single cross section as the basic unit. Taking the cross sections identified along the horizontal direction as an example: an identified cross section $[P_i, P_j]$ at row r is connected to another identified cross section $[P'_i, P'_j]$ at row $r + 1$ only when both the following two conditions are satisfied:

1. The two cross sections overlap along the vertical direction.
2. The difference in width between the two cross sections must be less than λ, a predefined threshold, which is set to 5 in the experiment.

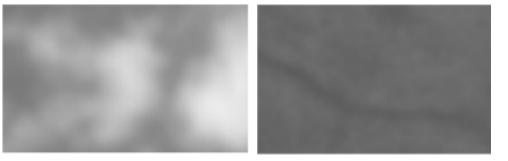


Fig. 4. Illustration of continuity: (a) the background region between the two bright lesions has two strong parallel, but short boundaries; (b) a less contrasted vessel that has an extended continuous linear structure.

In this way, the cross sections are connected to form potential candidates of vessel segment. Each potential vessel segment is assigned with a confidence value $C2$ that is determined by the two factors: the total number of cross sections in the segment (L) and the width variance of cross sections (σ) in the segment, as empirically defined in the following formula:

$$C2 = e^{-\alpha_2 * \sigma} \ln(L) \tag{2}$$

where α_2 is a positive weighting factor, set to 0.3 in the experiment. Higher confidence values are assigned to vessel segments with a larger number of cross sections and a smaller width variance. After assigning all pixels on a vessel segment with their corresponding $C2$ values and the pixels not on any vessel segment with the value 0, a 2D confidence map C_h is obtained to quantitatively measure the continuity of cross sections in the horizontal direction. Similarly, A 2D confidence map C_v can be obtained by connecting the cross sections identified vertically. An illustration of C_h and C_v is shown in Fig. 3 (c) and (d), where it is clear that this measurement effectively incorporates the contextual information for identifying vessels.

2.3 Bayesian Decision and Post Processing

There are four quantitative measurements (S_h, S_v, C_h, C_v) associated with each potential vessel segment. Bayesian rule is introduced here to statistically combine the four measurements and provide one confidence map for each potential vessel segment. With the horizontal measurements (S_h, C_h), the probability of being vessel given S_h and C_h can be computed with Bayesian rule as following:

$$P(Vessel|S_h, C_h) = \frac{P(Vessel)P(S_h, C_h|Vessel)}{P(S_h, C_h)} \tag{3}$$

Note that all pixels on a vessel segment have the same C_h value, but may have different S_h values. So the averaged S_h of each vessel segment is used in the equation (3). To make the problem attackable, we assume S_h and C_h are independent of each other and $P(Vessel)$ is a uniform distribution. Therefore, the equation (3) is simplified as

$$P(Vessel|S_h, C_h) \propto P(S_h|Vessel)P(C_h|Vessel) \tag{4}$$

Thus, the confidence map M_h is obtained by multiplying S_h and C_h element by element. Similarly, M_v, the confidence map for the vertical direction can also be obtained by multiplying the 2D map S_v and C_v. We define the final confidence map $M = max(M_h, M_v)$. Fig. 3 (e) illustrates the final result obtained using this method.

Sudden changes in vessel width at bifurcations cause the vessel segment connecting process described in Section 2.2 to fail. To connect different vessel segments at bifurcations, a simple method for region filling is applied. First, all regions in binary mask B are set to zero, where their corresponding values in M are not zero. For the remaining small isolated regions in B, if one region can connect multiple regions that correspond non-zero values in M, the region is then used for region filling and the corresponding M value is defined as the minimum in M. The non-maxima suppression introduced in [15] is recursively applied to the vessel boundary as a post processing step to reduce the boundary widening effect caused by Gaussian filtering in the Canny edge detection.

3 Experimental Results

The STARE data set [12] is used to evaluate the proposed algorithm and to facilitate performance comparison against several published papers using the same data set. Hand-labeled vessels as ground truth of 20 color retinal images are included in the data set. The images are with a size of 605×700 pixels in RGB format.

The performance of the proposed vessel detection method is measured using the Receiver Operating Characteristic (ROC) curve described in [12]. ROC curves reported in [12] and [15] are duplicated in the paper for the purpose of comparison. A better result reported in [13] is not included, since the method is based on a high-dimensional feature vector with supervised learning that requires extensive training samples. The comparison of ROC curves of the proposed algorithm and those in [12], [15] is shown in Fig. 5. The result shows that the proposed algorithm achieves a better performance on the vessel detection with the STARE data set. Visual inspection indicates that the results with the false positive rate higher than 0.04 are too noisy to be useful in the applications like width measurement. Two examples of the vessel detection results, one with lesions and the other without lesions, are also shown in Fig. 5. The existence of lesions makes the vessel detection more difficult, as seen in the figure.

As discussed in Section 1.2, the accurate detection is one of the goals of the proposed algorithm. Since no vessel detection method is perfect, another goal that may be more important for vessel diameter measurement is to provide a measure for the detection reliability of each vessel segment. The confidence map in the proposed method provides such a measurement that may have potential to guide the vessel selection process. As one example shown in Fig. 3 (e), the vessel segments that are more reliably detected have higher values in the confidence map.

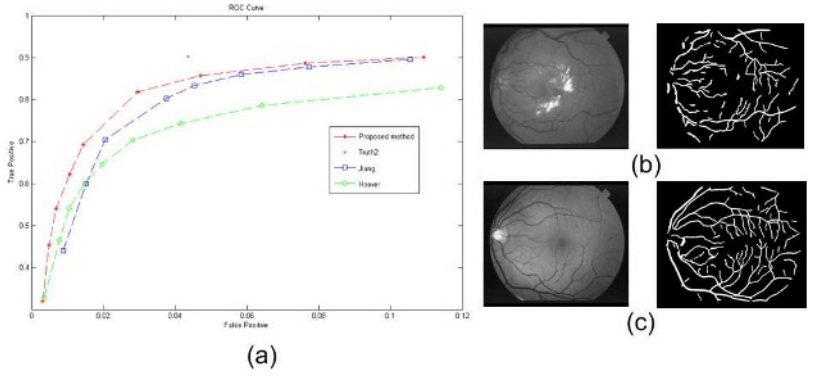


Fig. 5. Experimental results: (a) ROC curves with the STARE data set: curve with legend "proposed method" is obtained with the method proposed in this paper; curve with legend "Hoover" is duplicated from [12]; curve with legend "Jiang" is duplicated from [15]; the cross point with legend "Truth2" is obtained with another manually labeled result of the STARE data, used as the performance bound for the computational methods [12, 15]; (b) A result from a retinal image with bright lesions; (c) A result from a healthy retinal image.

4 Summary and Future Research

This paper explores the appearance and shape properties of vessel for retinal vessel detection, and a better detection accuracy is achieved. As a by-product, the proposed method provides a confidence value describing how reliable each detected vessel segment is detected.

There are other interesting topics of retinal vessel detection that are not discussed in this paper, but are ongoing research topics, for example, (1) sensitivity analysis of the parameters used; (2) extension of the method to the supervised learning case, where manually labelled vessel maps are available for training. In this setting, a Bayesian formula for computing likelihood can be obtained. (3) Experiments on larger data set of retinal images and a comprehensive evaluation of the method.

References

1. Wong, T., Klein, R., et al.: Retinal microvascular abnormalities and their relationship with hypertension, cardiovascular disease, and mortality. Survey of Ophthalmology **46**(1) (2001) 59–80
2. Mitchell, P., Leung, H., et al.: Retinal vessel diameter and open-angle glaucoma: the blue mountains eye study. Ophthalmology **112**(2) (2005) 245–250
3. Wong, T., Mitchell, P.: Hypertensive retinopathy. The New England Journal of Medicine **351**(22) (2004) 2310–2317
4. Klein, R., Klein, B., et al.: The relation of retinal vessel caliber to the incidence and progression of diabetic retinopathy: Xix: the wisconsin epidemiologic study of diabetic retinopathy. Archives of Ophthalmology **122**(1) (2004) 76–83

5. Wong, T., Rosamond, W., et al.: Retinopathy and risk of congestive heart failure. Journal of the American Medical Association **293**(1) (2005) 63–69
6. Hubbard, L., Brothers, R., et al.: Methods for evaluation of retinal microvascular abnormalities associated with hypertension/sclerosis in the atherosclerosis risk in communities study. Ophthalmology **106**(12) (1999) 2269–2280
7. Shen, H., Stewart, C., et al.: Frame-rate spatial referencing based on invariant indexing and alignment with application to online retinal image registration. IEEE Trans. on Pattern Analysis and Machine Intelligence **25**(3) (2003) 379–384
8. Foracchia, M., Grisan, E., et al.: Detection of optic disc in retinal images by means of a geometrical model of vessel structure. IEEE Trans. on Medical Imaging **23**(10) (2004) 1189–1195
9. Huang, K., Yan, M.: A local adaptive algorithm for microaneurysms detection in digital fundus images. In: ICCV Workshop of Computer Vision for Biomedical Image Applications. (2005)
10. Sofka, M., Stewart, C.: Retinal vessel extraction using multiscale matched filters, confidence and edge measures. Technical report, Department of Computer Science, Rensselaer Polytechnic Institute, NY, USA (2005)
11. Wu, D., Zhang, M., et al.: On the adaptive detection of blood vessels in retinal images. Technical report, Department of Computer Science, Texas A&M University, TX, USA (2005)
12. Hoover, A., Kouznetsova, V., et al.: Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE Trans. on Medical Imaging **19**(3) (2000) 203–210
13. Stall, J., Abramoff, M., et al.: Ridge-based vessel segmentation in color images of the retina. IEEE Trans. on Medical Imaging **23**(3) (2004) 501–509
14. Can, A., Shen, H., et al.: Rapid automated tracing and feature extraction from retinal fundus images using direct exploratory algorithms. IEEE Trans. on Information Technology in Biomedicine **3**(2) (1999) 125–138
15. Jiang, X., Mojon, D.: Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images. IEEE Trans. on Pattern Analysis and Machine Intelligence **25**(1) (2003) 131–137

Carotid Artery and Jugular Vein Tracking and Differentiation Using Spatiotemporal Analysis

David Wang[1,2], Roberta Klatzky[1], Nikhil Amesur[2], and George Stetten[1,2]

[1] Carnegie Mellon University, Pittsburgh, PA, USA
[2] University of Pittsburgh Medical Center, Pittsburgh, PA, USA

Abstract. We have derived and evaluated parameters from ultrasound images of the neck to permit a computer to automatically characterize and differentiate between the carotid artery and jugular vein at image acquisition time during vascular interventions, given manually placed seed points. Our goal is to prevent inadvertent damage to the carotid artery when targeting the jugular vein for catheterization. We used a portable 10 MHz ultrasound system to acquire cross sectional B-mode ultrasound images of these great vessels at 10 fps. An expert user identified the vessels in the first frame by touching the vessels on the screen with his fingertip, and the computer automatically tracked the vessels and calculated a best-fit ellipse for each vessel in each subsequent frame. Vessel location and radii were further analyzed to produce parameters that proved useful for differentiating between the carotid artery and jugular vein. These parameters include relative location of the vessels, distension of the vessel walls, and consistent phase difference between the arterial and venous pulsations as determined by temporal Fourier analysis.

1 Introduction

The common carotid artery (CCA) and the internal jugular vein (IJV) run side-by-side in the neck, one pair on the left and one on the right. The CCA carries oxygenated blood up to the head while the IJV drains deoxygenated blood down to the heart. The right IJV is a common entry site for intravascular procedures such as central catheter line placements, hemodynamic measurements, myocardial biopsies, cardiac ablations, etc. The right IJV is chosen because its path to the right atrium is straighter than that from the left IJV. Inadvertent CCA puncture while targeting the IJV is reported to have an incidence of 2-8% [1, 2] and usually results in localized hematoma. The hematoma may enlarge rapidly if the patient is coagulopathic, or if a large puncture wound is produced by the introduction of the sheath itself into the CCA. Potential negative consequences of CCA puncture include airway obstruction [3], pseudoaneurysm [4], arterio-venous fistula formation [5], retrograde aortic dissection [6], and cerebrovascular accident [7].

The introduction of B-mode ultrasound to guide IJV access has decreased the arterial puncture incidence [8]. The direction and pattern of flow from color Doppler can further help distinguish arteries from veins, but since doctors generally hold the ultrasound transducer perpendicular to the target vessels during vascular access, any slight

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 654 – 661, 2006.

angular deviation may reverse the perceived flow and make identification of the vessels ambiguous. Thus color Doppler is generally only used during pre-operative evaluation and not during the actual procedure.

We believe a semi-automated vessel identification system operating in real time during the procedure could further decrease the incidence of inadvertent CCA puncture and expedite cannulation. We expect this system to be used in conjunction with a new method of ultrasound display, the Sonic Flashlight [9], which employs a half-silvered mirror to project the image directly within the patient. By analyzing the properties of the tracked objects, it should be possible to automatically classify and graphically mark each as an artery, vein, or other tissue type. This could help prevent the operator, especially the relative novice, from accidentally puncturing the CCA.

Due in part to the presence of speckle, ultrasound is among the noisiest of all medical imaging modalities, and its anisotropy and path dependence add further difficulties for image analysis. Many researchers have developed ultrasound tracking techniques. Abolmaesumi, et al., described methods to automatically track and segment the boundary of the CCA by using the "Star" algorithm, stabilized by a temporal Kalman filter [10]. However, it requires initialization using prior knowledge about probe movement. Yeung, et al., showed that objects in sequential ultrasound images can be tracked using a deformable mesh model [11], but not in real-time. Nakayama, et al., designed an ultrasonic method of measuring arterial wall movement [12], but it requires manual outlining of the vessels and scanning in longitudinal section rather than cross section. Methods to identify anatomical targets other than vessel lumens have also been investigated. Drukker, et al., researched the use of a radial gradient index filter to automatically detect lesions in breast ultrasound [13]. Ladak, et al., developed automatic methods to segment the boundaries of the prostate [14]. Algorithms also exist for automatically locating instruments such as biopsy needles in an ultrasound image [15]. Wilson described methods using Doppler to automatically recognize the walls of a blood vessel and determine its orientation and diameter [16].

We propose to solve the problem of identifying the IJV and CCA by creating a system that will simultaneously track the vessels and provide statistical data for their identification. Our current emphasis is on identifying which of the parameters derived from our tracking algorithms best identify and distinguish between the two vessels.

We start in Section 2.1 by introducing the *spokes ellipse* algorithm for tracking and calculating the cross sectional elliptical radii of each vessel. In Sections 2.2 and 2.3 we explain how we calculated our parameters for differentiating the IJV from the CCA. In Section 2.4 we describe how we collected data from 16 healthy volunteers. In Section 2.5 we explain how we validated the *spokes ellipse* algorithm. In Section 3 we discuss the results. In Section 4 we conclude with directions for future work.

2 Methods

2.1 *Spokes Ellipse* Algorithm

We have developed an efficient algorithm that tracks blood vessels in real time and simultaneously calculates their elliptical radii and cross sectional area in each frame to be used in further analysis. An expert simply taps each vessel once on the screen,

and the algorithm outlines and tracks the vessel walls in subsequent frames (Figure 1). Much like the above-mentioned "Star" algorithm, it initially draws radial lines emanating from the seed point. An intensity-based boundary detection algorithm searches for the most likely boundary along each spoke, from which an ellipse is fitted by a least squares method [17]. The cross sectional area of the vessel is approximated by the area of the ellipse. The center of the ellipse is then used as the seed point for the spokes in the next frame. By recalculating the center of the vessel and its boundaries in each frame, the vessel can be tracked in real-time, although sudden movement of the transducer may cause the tracking to be lost. This algorithm is run twice in each frame – once for the CCA and once for the IJV.

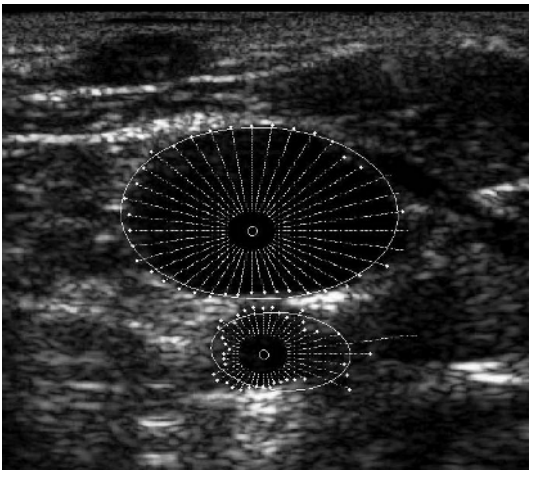


Fig. 1. *Spokes ellipse* algorithm applied to the IJV (top) and the CCA (bottom). Spokes grow until they reach a boundary (white dot) or a pre-set maximum length. Ellipses are fit to the dots for each vessel. Algorithm runs in real-time.

2.2 Frequency and Phase Calculations

By taking the Fourier transform of the series of cross sectional areas of each vessel, we can obtain the magnitude and phase at each frequency. The heart rate can be determined by identifying the frequency with the greatest magnitude between 40 and 150 cycles per minute (cpm). Similarly, the respiratory rate can be determined as the frequency with greatest magnitude in the range 10 to 30 cpm. Recall that the Fourier transform $X(\omega)$ of a signal $x(t)$ can be represented in phasor notation as

$$X(\omega) = r(\omega)e^{j\theta(\omega)} \tag{1}$$

where $r(\omega) \geq 0$ is the magnitude and $-\pi < \theta(\omega) \leq \pi$ is the phase in the complex plane. Denoting the Fourier transforms of the IJV and CCA cross-sectional areas respectively as $J(\omega)$ and $C(\omega)$, the phase difference $Q(\omega)$ between the CCA and IJV can be found from their ratio, normalized by their relative magnitudes,

$$Q(\omega) = \frac{r_J(\omega)}{r_C(\omega)} \frac{C(\omega)}{J(\omega)} = e^{j(\theta_C(\omega) - \theta_J(\omega))}. \tag{2}$$

This yields a unit phasor, the phase of which is the difference between the phases of the CCA and IJV at frequency ω,

$$\Delta\theta(\omega) = \theta_C(\omega) - \theta_J(\omega). \tag{3}$$

Thus,

$$\Delta\theta(\omega) = \arctan \frac{\mathrm{Im}\{Q(\omega)\}}{\mathrm{Re}\{Q(\omega)\}} \tag{4}$$

Since we have a sampled Fourier transform, individual phase samples may or may not represent a consistent phase difference, so we convolve over a narrow band in the frequency domain with a normalized Gaussian smoothing filter $G(\omega, \sigma)$ as follows

$$\Delta\tilde{\theta}(\omega) = \arctan \frac{\mathrm{Im}\{G(\omega,\sigma) * Q(\omega)\}}{\mathrm{Re}\{G(\omega,\sigma) * Q(\omega)\}} \tag{5}$$

With σ set empirically to 3 samples or 2.5 cpm in the Fourier frequency domain. (Note that convolution can be applied to the complex number $Q(\omega)$ by applying it independently to the real and imaginary parts). A measure of phase consistency for this smoothed phase difference is

$$\alpha(\omega) = \left| G(\omega,\sigma) * Q(\omega) \right|. \tag{6}$$

Random phase differences in individual samples of the Fourier transforms due to noise tend to cancel upon convolution with the Gaussian, yielding a value of $\alpha(\omega)$ near 0, whereas a consistent phase shift over consecutive frequency samples of the Fourier transforms yields a value of $\alpha(\omega)$ near 1.

2.3 Other Differentiating Factors

Besides phase, other measures might differentiate the IJV from the CCA. In general, an artery has a thicker muscular layer than a vein. Thus the artery is stiffer while the vein is more compliant. A stiffer material emphasizes high frequencies and attenuates low frequencies. A compliant material does the opposite. We analyzed this phenomenon by calculating the slope of the linear regression that best fit the Fourier transforms from 10-250 cpm.

Since the vein is more compliant and contains blood at a lower pressure, it is less likely to assume a circular shape than an artery. We calculated the eccentricity of the two vessels as

$$eccentricity = \sqrt{1 - \frac{b^2}{a^2}} \tag{7}$$

where a is the semi-major axis and b is the semi-minor axis. We also looked at whether the IJV is consistently closer to the skin than the CCA.

2.4 Experimental Protocol

We recruited 16 healthy volunteers ages 20 to 51. The volunteers were asked to lie supine with their heads turned to the left. The right sides of their necks were scanned with ultrasound in a standard manner. A series of digitized images was recorded from the ultrasound machine at 10 fps for 120 seconds. Heart rates were measured by palpation before and after the scan, and the average used to estimate heart rate during the scan. Based on a preliminary data set, our exclusion criteria were IJV eccentricity > 0.8, CCA eccentricity > 0.6, or RMS ellipse fitting error > 0.02 for either vessel. Excluded frames generally contained vessels that departed substantially from an elliptical form. This constituted approximately 15% of the ellipses that were fit, representing ultrasound transducer locations and orientations that would likely be deemed unsuitable for cannulation with or without our algorithm in the clinical setting. The *spokes ellipse* algorithm was applied to all frames collected from each subject, but only the first 500 continuous frames satisfying these inclusion criteria were used in our analysis.

2.5 Validation of the *Spokes Ellipse* Algorithm

The *spokes ellipse* algorithm requires two assumptions: (1) the vessel cross sections are elliptical, and (2) the algorithm-drawn ellipses are similar to manual traces of the lumen drawn by an expert. We define the similarity between two areas as

$$\frac{Area\ 1 \cap Area\ 2}{Area\ 1 \cup Area\ 2} \times 100 \quad \text{(the ratio expressed as a percentage)} \tag{8}$$

A perfect similarity would thus have a value of 100%. We randomly sampled 100 images from the 8000 frames (500 frames per subject, 16 subjects). An expert outlined the lumens of both the CCA and IJV in all 100 images. We then tested the similarity between the area of the ellipse fit to the expert tracing and (1) the expert traced area itself, and (2) the algorithm defined ellipse, to test assumptions 1 and 2, respectively.

3 Results

Among the 100 sampled images, there was a 91±1% similarity between the manual tracings of the lumen and the ellipses fitted to those tracings, suggesting that the cross-sectional areas of the vessels are reasonably elliptical (assumption 1). The similarity between ellipses fit to the manual tracings and the ellipses fit by the *spokes-ellipse* algorithm was 82±2% for IJV and 79±2% for CCA. Moreover, the areas of the ellipses fit to the manual tracings and algorithm-fit ellipses showed a correlation value of r = .80. This suggests that while the algorithm may not have fully extracted the elliptical shape defined by an expert (assumption 2), it tended to report area with a consistent bias that preserved the relative sizes of the CCA and IJV and their individual changes over time.

An expert user verified that the algorithm successfully tracked the vessels in the 500 frames that satisfied our inclusion criteria. These frames were subsequently used

to calculate the parameters described in Sections 2.2 and 2.3 to test their reliability at differentiating between the CCA and IJV. The results are shown in Table 1.

The heart rates calculated from the CCA were not significantly different from those from the IJV. The respiratory rates also did not differ significantly. This is to be expected, since the underlying heart and respiratory rates are identical for both vessels. The fact that these heart rates are consistent with those determined by palpation during the scan (data not shown) lends further validity to our approach.

The slopes of the Fourier transform of the IJV (-36.99±14.14) were significantly more negative than those of the CCA (-16.36±2.58), with a p-value < 0.05. This is consistent with the IJV being more compliant than the CCA and thus attenuating the higher frequencies.

Table 1. 95% confidence interval for 8 parameters, derived separately (first 5 parameters) or jointly (last 3 parameters) from the CCA and IJV data. The last column shows, for the first 5 parameters, the significance level (*p*) for a paired t-test comparing the CCA and IJV values across the sampled frames.

Parameter	Vessel	95% Confidence Interval	*p*-val.
Heart rate (cpm)	CCA	61.06 ± 6.79	0.40
	IJV	60.91 ± 6.88	
Respiratory rate (cpm)	CCA	15.81 ± 1.80	0.16
	IJV	17.12 ± 2.26	
Slope of Fourier transform	CCA	-16.36 ± 2.58	0.006
	IJV	-36.99 ± 14.14	
Mean of eccentricity	CCA	0.62 ± 0.04	0.42
	IJV	0.63 ± 0.06	
Standard deviation of eccentricity	CCA	0.11 ± 0.02	0.006
	IJV	0.08 ± 0.02	
Smallest ratio of distance to skin from CCA vs. IJV		1.35 ± 0.07	N/A
Phase difference between CCA and IJV at heart rate		-1.95 ± 0.84	N/A
Phase consistency between CCA and IJV at heart rate		0.71 ± 0.10	N/A

It was expected that, being more compliant, the IJV would show a greater degree of eccentricity, especially considering the pressure being applied by the ultrasound transducer. However, the respective means of eccentricity for the CCA and IJV were 0.62±0.04 and 0.63±0.06, a non-significant difference, indicating that the eccentricity of the vessels is not a good distinguishing feature. Eccentricity is also very dependent on the angle between image plane and direction of the vessels, so even circular vessels can appear eccentric when scanned at an angle. The slightly higher standard deviation of eccentricity for the CCA also conflicts with the expectation that the IJV would show a greater variation in eccentricity over time, and as such is suspect as a viable differentiating factor.

It was expected that the CCA would lie deeper on average than the IJV. Thus the ratio of the distance to the skin from the center of the CCA to that of the IJV should be greater than 1. Indeed, even the smallest ratios in all frames for each subject show a 95% confidence interval that excludes 1.0. However, while this observation is true when scanning at the “conventional” site, it may no longer be true when scanning at

somewhat displaced sites on the neck. One study shows the IJV is positioned completely lateral to the CCA on the image display instead of overlapping the CCA in 8.7% of clinical scans [18]. Doctors often prefer such laterally placed IJV because CCA puncture is less likely when it is not directly under the IJV. Thus, doctors may seek scanning positions that render the vessel depth parameter invalid.

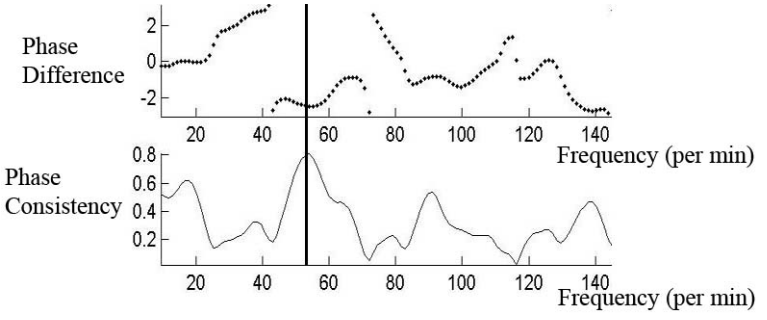


Fig. 2. (Top) Phase difference between the CCA and the IJV for subject 4 whose heart rate is around 55/min (bold vertical line). (Bottom) Phase consistency measure (see Section 2.2).

The phase difference at the heart rate has a 95% confidence interval of -1.95±0.84 radians. If we had reversed the CCA and IJV in the calculation of phase difference, the 95% confidence interval would be just the negative, i.e., 1.95±0.84 radians. Since these intervals do not include 0 nor π, it appears that phase difference can be unambiguously used to differentiate between the CCA and the IJV. In addition, phase difference shows high consistency around the frequency of the heart rate. For example, Figure 2 shows the phase (top) and consistency (bottom) for patient 4. In the region of the heart rate (55 cpm) the consistency approaches 1, lending confidence to the recorded phase difference of approximately -2 radians in this region of the spectrum.

4 Conclusions and Future Work

We have shown in each of our 16 subjects, that the spokes ellipse algorithm could maintain tracking for at least 50 seconds (500 continuous frames) during the 120 second ultrasound scans. Although the CCA and IJV both pulsate, we have verified that the slope of the Fourier transform, the vertical locations of the vessels, and phase difference can be used individually to reliably distinguish between the vessels.

We have yet to determine what combinations of parameter weights would further optimize differentiation. Also needed is a method to reliably detect loss of tracking. While the *spokes ellipse* algorithm runs in real-time, the vessel classification algorithm at present incurs a lag time because the Fourier transform requires a significant number of cardiac cycles to operate. We are considering various other phase-detection methods besides Fourier that can operate within a single cardiac cycle. Future work may also incorporate color Doppler to initially identify the blood vessels, replacing manual initialization. We plan to obtain images from patients who are about to

undergo IJV catheter insertions in the hospital to more accurately reflect the parameters in the clinical setting. Other smaller vessels in the vicinity of the CCA and IJV will also need to be accounted for. Eventually we plan to incorporate our automated vessel identification system into ongoing clinical trials of the Sonic Flashlight.

References

1. Davies MJ, Cronin KD, and Domaingue CM, "Pulmonary artery catheterization: an assessment of risks and benefits in 220 surgical patients," *Anaesth Intensive Care* 10:9, 1982.
2. Patel C, Laboy V, Venus B, Mathru M, Wier D, "Acute complications of pulmonary artery catheter insertion in critically ill patients," *Crit Care Med*, 14(3):195-7 1986 Mar
3. Kua JS, Tan IK, et al., "Airway obstruction following internal jugular vein cannulation," *Anaesthesia*, 52:776-80, 1997
4. Aoki H., Mizobe T., Nozuchi S., et al, "Vertebral artery pseudoaneurysm: A rare complication of internal jugular vein catheterization," *Anesth Analg* 75:296-298,1992
5. Gobeil F. Couture P. Girard D, Plante R, "Carotid Artery-lnternal Jugular Fistula: Another Complication following Pulmonary Artery Catheterization via the Internal Jugular Venous Route," *Anesthesiology* 1994, 80:23-232.
6. Applebaum RM, et al, "Transesophageal echocardiographic identification of a retrograde dissection of the ascending aorta caused by inadvertent cannulation of the common carotid artery," *J Am Soc Echocardiogr* 1997 Sep;10(7):749-51
7. Zaidi NA, Khan M, Naqvi HI, Kamal RS, "Cerebral infarct following central venous cannulation," *Anaesthesia*, 53:186-91, 1998 Feb
8. Denys BG, et al, "Ultrasound-assisted cannulation of the internal jugular vein: A prospective comparison to the external landmark-guided technique," *Circulation* 1993; 87:1557-62.
9. Stetten GD, Chib V, "Overlaying Ultrasound Images on Direct Vision," *J Ultrasound Med* 2001;20(3):235-40.
10. Abolmaesumi P, et al., "Real-Time Extraction of Carotid Artery Contours from Ultrasound Images" 13th IEEE Symposium on Computer-Based Medical Systems (CBMS'00) p. 181
11. Yeung F, et al., "Feature-Adaptive Motion Tracking of Ultrasound Image Sequences Using A Deformable Mesh" IEEE Transactions on Medical Imaging Vol. 17, No. 6, Dec 1998
12. Nakayama K, Sato S, "Ultrasonic measurement of arterial wall movement utilizing phase-tracking systems," *Proceedings of the 10th International Congress on Medical and Biological Engineering*, Dresden, Germany, 1973:318. Abstract
13. Drukker K, et al., "Computerized lesion detection on breast ultrasound," *Medical Physics*, July 2002, Volume 29, Issue 7, pp. 1438-1446
14. Ladak H, et al., "Prostate boundary segmentation from 2D ultrasound images" *Medical Physics*, August 2000, Volume 27, Issue 8, pp. 1777-1788
15. Draper K, et al., "An algorithm for automatic needle localization in ultrasound-guided breast biopsies," *Medical Physics*, August 2000, Volume 27, Issue 8, pp. 1971-1979
16. Wilson LS, Dadd MJ, Gill RW, "Automatic vessel tracking and measurement for Doppler studies," *Ultrasound Med Biol.* 1990;16(7):645-52
17. Pilu M, Fitzgibbon A, Fisher R, "Ellipse-specific direct least-square fitting," *Proceedings of the IEEE international Conference on Image Processing*, IEEE Computer Society Press, Los Alamitos, CA 1996; 3:599-602.
18. Troianos CA, et al., "Internal Jugular Vein and Carotid Artery Anatomic Relation as Determined by Ultrasonography," *Anesthesiology* 1996; 85:43-8

Appearance Models for Robust Segmentation of Pulmonary Nodules in 3D LDCT Chest Images

Aly A. Farag[1], Ayman El-Baz[1], Georgy Gimel'farb[2], Robert Falk[3], Mohamed A. El-Ghar[4], Tarek Eldiasty[4], and Salwa Elshazly[1]

[1] Computer Vision and Image Processing Laboratory, University of Louisville, Louisville, KY 40292
{farag, elbaz}@cvip.Louisville.edu
http://www.cvip.louisville.edu
[2] Department of Computer Science, Tamaki Campus, University of Auckland, Auckland, New Zealand
[3] Director, Medical Imaging Division, Jewish Hospital
[4] Urology and Nephrology Department, University of Mansoura, Mansoura, Egypt

Abstract. To more accurately separate each pulmonary nodule from its background in a low dose computer tomography (LDCT) chest image, two new adaptive probability models of visual appearance of small 2D and large 3D pulmonary nodules are used to control evolution of deformable boundaries. The appearance prior is modeled with a translation and rotation invariant Markov-Gibbs random field of voxel intensities with pairwise interaction analytically identified from a set of training nodules. Appearance of the nodules and their background in a current multi-modal chest image is also represented with a marginal probability distribution of voxel intensities. The nodule appearance model is isolated from the mixed distribution using its close approximation with a linear combination of discrete Gaussians. Experiments with real LDCT chest images confirm high accuracy of the proposed approach.

1 Introduction

Because lung cancer is the most common cause of cancer deaths, fast and accurate analysis of pulmonary nodules is of major importance for medical computer-aided diagnostic systems (CAD). In [1] we introduced a fully automatic nodule detection algorithm showing the accuracy up to 93.3% on the experimental database containing 200 real LDCT chest data sets with 36,000 2D slices. Below we focus in the next CAD stage, namely, on accurate segmentation of the detected nodules for subsequent volumetric measurements to monitor how the nodules change in time.

We use a two-step procedure to separate the nodules from their background: (*i*) an initial LDCT slice is segmented with algorithms introduced in [2] to isolate lung tissues from surrounding structures in the chest cavity as shown in Fig. 1 and (*ii*) the nodules in the isolated lung regions are segmented by evolving deformable boundaries under forces that depend on the learned current and prior

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 662–670, 2006.

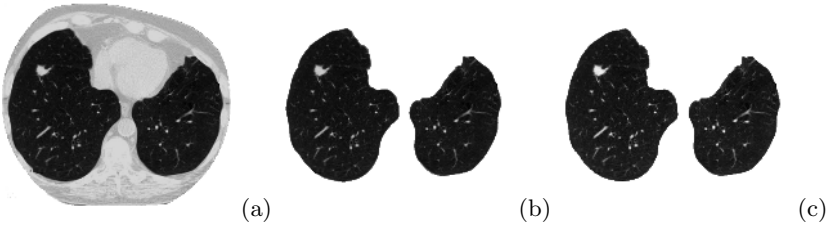


Fig. 1. Step 1 of our segmentation approach (e.g., [2]): an LDCT slice (a) with isolated lungs (b), and (c) the normalized segmented lung image

appearance models. At Step 1 each LDCT slice is modelled as a bi-modal sample from a simple Markov–Gibbs random field of interdependent region labels and conditionally independent voxel intensities (gray levels). This step is necessary for more accurate separation of nodules from the lung tissues at Step 2 because voxels of both the nodules and other chest structures around the lungs are of quite similar intensity.

Previous work. At present, segmentation of pulmonary nodules is under extensive studies. Typical conventional techniques are based on fitting a Gaussian model to empirical data [3] but this approach becomes a challenge when initial measurements are corrupted with outliers and margin-truncation due to neighboring structures. Okada et al. [4] proposed an anisotropic intensity model fitting with analytical parameter estimation. Zhao et al. [5] and Kostis et al. [6] proposed to segment 2D and 3D nodules based on thresholding the voxel intensity. Their algorithms accurately segment well-defined solid nodules with similar average intensities but become unreliable on the cavity or non-solid nodules. ***Our segmentation overcomes these drawbacks due to using at Step 2 deformable boundary models such that their evolution is controlled by both a learned prior probability model of the visual nodule appearance and an adaptive appearance model of the nodules in a current image to be segmented.***

Basic notation. Let (x, y, z) denote Cartesian coordinates of points in a finite arithmetic lattice $\mathbf{R} = [(x, y, z) : x = 0, \ldots, X-1; y = 0, \ldots, Y-1, z = 1, \ldots, Z-1]$. It supports a given 3D grayscale image $\mathbf{g} = [g_{x,y,z} : (x, y, z) \in \mathbf{R}; g_{x,y,z} \in \mathbf{Q}]$ with gray levels from a finite set $\mathbf{Q} = \{0, \ldots, Q-1\}$ and its region map $\mathbf{m} = [m_{x,y,z} : (x, y, z) \in \mathbf{R}; m_{x,y,z} \in \mathbf{L}]$ with region labels from a finite set $\mathbf{L} = \{\mathsf{nd}, \mathsf{bg}\}$. Each label $m_{x,y,z}$ indicates whether the pixel (x, y, z) in the corresponding data set $\mathbf{g}$ belongs to the goal object (pulmonary nodule), $m_{x,y,z} = \mathsf{nd}$, or to the background, $m_{x,y,z} = \mathsf{bg}$. Let $\mathbf{b} = [\mathbf{P}_k : k = 1, \ldots, K]$ be a deformable piecewise-linear boundary with K control points $\mathbf{P}_k = (x_k, y_k, z_k)$. The index k can be considered as a real number in the interval $\mathbf{K}$ indicating continuous positions around the boundary, e.g. $\mathbf{K} = [1, K]$ for the positions from $\mathbf{P}_1$ to $\mathbf{P}_K$.

Conventional deformable model. moves in the direction that minimizes a boundary energy E such as e.g. in [7]:

$$E = E_{\text{int}} + E_{\text{ext}} = \int_{k \in \mathbf{K}} \left(\xi_{\text{int}} \left(\mathbf{b}(\mathbf{P}_k) \right) + \xi_{\text{ext}} \left(\mathbf{b}(\mathbf{P}_k) \right) \right) dk \tag{1}$$

where $\xi_{\text{int}} \left(\mathbf{b}(\mathbf{P}_k) \right)$ and $\xi_{\text{ext}} \left(\mathbf{b}(\mathbf{P}_k) \right)$ are internal and external forces, respectively.

In this paper we present a new class of the the external energy that guided the evolution of deformable model based on two new probability models that roughly describe the prior and current visual appearance of the nodules.

2 Data Normalization

To account for monotone (order-preserving) changes of signals (e.g. due to different illumination or sensor characteristics), for each segmented data set, we will calculate the occurrence histogram, then we normalize the segmented data set to make $q_{max} = 255$ for each segmented data set, (e.g see Fig. 1(c)).

3 MGRF-Based Prior Appearance Model

To exclude an alignment stage before segmentation, the appearance of both small 2D and large 3D nodules is modeled with a translation and rotation invariant generic MGRF with voxel-wise and central-symmetric pairwise voxel interaction specified by a set $\mathbf{N}$ of characteristic central-symmetric voxel neighborhoods $\{\mathbf{n}_\nu : \nu \in \mathbf{N}\}$ on $\mathbf{R}$ and a corresponding set $\mathbf{V}$ of Gibbs potentials, one potential per neighborhood.

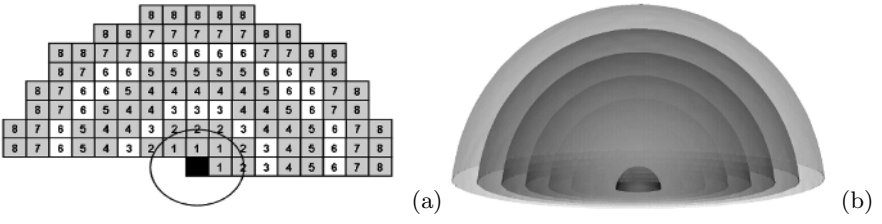


Fig. 2. Central-symmetric 2D (a) and 3D (b) neighborhoods for the eight distance ranges $[d_{\nu,\min} = \nu - 0.5, d_{\nu,\max} = \nu + 0.5)$; $\nu \in \mathbf{N} = \{1, \ldots, 8\}$ on the lattice $\mathbf{R}$

A central-symmetric voxel neighborhood $\mathbf{n}_\nu$ embraces all voxel pairs such that (x, y, z)-coordinate offsets between a voxel (x, y, z) and its neighbor (x', y', z') belong to an indexed semi-open interval $[d_{\nu,\min}, d_{\nu,\max})$; $\nu \in \mathbf{N} \subset \{1, 2, 3, \ldots\}$ of the inter-voxel distances: $d_{\nu,\min} \leq \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2} < d_{\nu,\max}$. Figure 2 illustrates the neighborhoods $\mathbf{n}_\nu$ for the uniform distance ranges $[\nu - 0.5, \nu + 0.5)$; $\nu \in \mathbf{N} = \{1, \ldots, 8\}$.

The interactions in each neighborhood $\mathbf{n}_\nu$ have the same Gibbs potential function $\mathbf{V}_\nu$ of gray level co-occurrences in the neighboring voxel pairs, and the voxel-wise interaction is given with the potential function $\mathbf{V}_{\text{vox}}$ of gray levels in the voxels:

$$\mathbf{V} = \left[\mathbf{V}_{\text{vox}} = [V_{\text{vox}}(q) : q \in \mathbf{Q}] ; \left\{ \mathbf{V}_\nu = \left[V_\nu(q, q') : (q, q') \in \mathbf{Q}^2 \right] : \nu \in \mathbf{N} \right\} \right]$$

Model identification. Let $\mathbf{R}_t = \{(x,y,z) : (x,y,z) \in \mathbf{R} \wedge m_{t;x,y,z} = \mathsf{nd}\}$ and $\mathbf{C}_{\nu,t}$ denote the part of the 3D lattice $\mathbf{R}$ supporting the training nodules in the image–map pair $(\mathbf{g}_t, \mathbf{m}_t) \in \mathbf{S}$ and the family of voxel pairs in $\mathbf{R}_t^2$ with the co-ordinate offsets $(\xi,\eta,\gamma) \in \mathbf{n}_\nu$, respectively. Let $\mathbf{F}_{\mathrm{vox},t}$ and $\mathbf{F}_{\nu,t}$ be a joint empirical probability distribution of gray levels and of gray level co-occurrences in the training nodules from the image $\mathbf{g}_t$, respectively: $\mathbf{F}_{\mathrm{vox},t} = [f_{\mathrm{vox},t}(q) = \frac{|\mathbf{R}_{t,q}|}{|\mathbf{R}_t|};\ \sum_{q\in\mathbf{Q}} f_{\mathrm{vox},t}(q) = 1]$ and $\mathbf{F}_{\nu,t} = [f_{\nu,t}(q,q') = \frac{|\mathbf{C}_{\nu,t;q,q'}|}{|\mathbf{C}_{\nu,t}|}; \sum_{(q,q')\in\mathbf{Q}^2} f_{\nu,\mathrm{t}}(q,q') = 1]$ where $\mathbf{R}_{t,q} = \{(x,y,z) : (x,y,z) \in \mathbf{R}_t \wedge g_{x,y,z} = q\}$ is a subset of voxels supporting the gray level q in the training nodules from the image $\mathbf{g}_t$ and $\mathbf{C}_{\nu,t;q,q'}$ is a subfamily of the voxel pairs $\mathbf{c}_{\xi,\eta,\gamma}(x,y,z) = ((x,y,z),(x+\xi,y+\eta,z+\gamma)) \in \mathbf{R}_t^2$ supporting the gray level co-occurrence (q,q') in the same nodules, respectively.

The MGRF model of the t-th object is specified by the joint Gibbs probability distribution on the sublattice $\mathbf{R}_t$:

$$P_t = \tfrac{1}{Z_t} \exp\left(|\mathbf{R}_t|\left(\mathbf{V}_{\mathrm{vox}}^{\mathsf{T}}\mathbf{F}_{\mathrm{vox},t} + \sum_{\nu\in\mathbf{N}} \rho_{\nu,t}\mathbf{V}_{\nu,t}^{\mathsf{T}}\mathbf{F}_{\nu,t}\right)\right) \tag{2}$$

where $\rho_{\nu,t} = |\mathbf{C}_{\nu,t}|/|\mathbf{R}_t|$ is the average cardinality of the neighborhood $\mathbf{n}_\nu$ with respect to the sublattice $\mathbf{R}_t$.

To simplify notation, let areas of the training nodules be similar, so that $|\mathbf{R}_t| \approx R_{\mathsf{nd}}$ and $|\mathbf{C}_{\nu,t}| \approx C_{\nu,\mathsf{nd}}$ for $t = 1,\ldots,T$, where R_{nd} and $C_{\nu,\mathsf{nd}}$ are the average cardinalities over the training set $\mathbf{S}$. Assuming the independent samples, the joint probability distribution of gray values for all the training nodules is as follows:

$$P_{\mathbf{S}} = \frac{1}{Z}\exp\left(TR_{\mathsf{nd}}\left(\mathbf{V}_{\mathrm{vox}}^{\mathsf{T}}\mathbf{F}_{\mathrm{vox}} + \sum_{\nu\in\mathbf{N}} \rho_\nu \mathbf{V}_\nu^{\mathsf{T}}\mathbf{F}_\nu\right)\right)$$

where $\rho_\nu = C_{\nu,\mathsf{nd}}/R_{\mathsf{nd}}$, and the marginal empirical distributions of gray levels $\mathbf{F}_{\mathrm{vox},\mathsf{nd}}$ and gray level co-occurrences $\mathbf{F}_{\nu,\mathsf{nd}}$ describe now all the nodules from the training set. Zero empirical probabilities caused by a relatively small volume of the training data available to identify the above model are eliminated if fractions defining the empirical probabilities in terms of cardinalities of the related sublattices or subfamilies are modified as follows: $(\langle\text{nominator}\rangle + \varepsilon)/(\langle\text{denominator}\rangle + S\varepsilon)$. With the Bayesian quadratic loss estimate, $\varepsilon = 1$ and $S = Q$ for the first-order or $S = Q^2$ for the second-order interactions.

Using the analytical approach similar to that in [2], the potentials are approximated with the scaled centered empirical probabilities:

$$\begin{aligned} V_{\mathrm{vox},\mathsf{nd}}(q) &= \lambda\left(f_{\mathrm{vox},\mathsf{nd}}(q) - \tfrac{1}{Q}\right);\quad (q) \in \mathbf{Q};\\ V_{\nu,\mathsf{nd}}(q,q') &= \lambda\left(f_{\nu,\mathsf{nd}}(q,q') - \tfrac{1}{Q^2}\right);\ (q,q') \in \mathbf{Q}^2; \nu \in \mathbf{N} \end{aligned} \tag{3}$$

where the common factor λ is also computed analytically. It can be omitted ($\lambda = 1$) if only relative potential values are used for computing relative energies $E_{\nu,\mathrm{rel}}$ of the central-symmetric pairwise voxel interactions in the training data. The energies that are equal to the variances of the co-occurrence distributions:

$$E_{\nu,\text{rel}} = \sum_{q,q' \in \mathbf{Q}^2} f_{\nu,\mathsf{nd}}(q,q') \left(f_{\nu,\mathsf{nd}}(q,q') - \frac{1}{Q^2} \right)$$

allow for ranking all the central-symmetric neighborhoods $\mathbf{n}_\nu$ and selecting the top-rank, i.e. most characteristic ones $\mathbf{N}' \subset \mathbf{N}$ to include to the prior appearance model of Eq. (3). Under the model, any grayscale pattern within a deformable boundary $\mathbf{b}$ in an image $\mathbf{g}$ is described by its Gibbs energy

$$E(\mathbf{g},\mathbf{b}) = \mathbf{V}^{\mathsf{T}}_{\text{vox},\mathsf{nd}} \mathbf{F}_{\text{vox},\mathsf{nd}}(\mathbf{g};\mathbf{b}) + \sum_{\nu \in \mathbf{N}'} \mathbf{V}^{\mathsf{T}}_{\nu,\mathsf{nd}} \mathbf{F}_{\nu,\mathsf{nd}}(\mathbf{g},\mathbf{b}) \tag{4}$$

where $\mathbf{N}'$ is an index subset of the selected top-rank neighborhoods, and the empirical probability distributions are collected within the boundary $\mathbf{b}$ in $\mathbf{g}$.

4 LCDG-Based Current Appearance Model

The visual appearance of nodules in each current data set $\mathbf{g}$ to be segmented typically differ from the appearance of the training nodules due to non-linear intensity variations from different data acquisition systems and changes in patient tissue characteristics, radiation dose, scanner type, and scanning parameters. This is why, in addition to the appearance prior learned from the normalized training nodules, we model the marginal gray level distribution within an evolving boundary $\mathbf{b}$ in $\mathbf{g}$ with a dynamic mixture of two distributions for current candidates for nodules and their background, respectively. The mixture is closely approximated with a bi-modal linear combination of discrete Gaussians (LCDG) and then partitioned into the nodule and background LCDGs. The approximation is performed with the modified EM-based approach in [2].

5 Boundary Evolution Using Two Appearance Models

The following external energy term in Eq. (1) combines the learned prior and current appearance models to guide an evolving boundary in a way such that maximizes the energy within the boundary:

$$\xi_{ext}\left(\mathbf{b}(\mathbf{P}_k = (x,y,z))\right) = -p_{\text{vox},\mathsf{nd}}(g_{x,y,z}) \pi_{\mathbf{P}}(g_{x,y,z}|\mathbf{S}) \tag{5}$$

where $p_{\text{vox},\mathsf{nd}}(q)$ is the marginal probability of the gray level q in the LCDG model for the nodules, arteries, and veins (see Section 4) and $\pi_{\mathbf{P}}(q|\mathbf{S})$ is the prior conditional probability of the gray level q, given the current gray values in the characteristic central-symmetric neighborhoods of $\mathbf{P}_k$, for the MGRF prior model in Section 3:

$$\pi_{\mathbf{P}}(g_{x,y,z}|\mathbf{S}) = \frac{\exp\left(E_{\mathbf{P}}(g_{x,y,z}|\mathbf{S})\right)}{\sum\limits_{q \in \mathbf{Q}} \exp\left(E_{\mathbf{P}}(q|\mathbf{S})\right)}$$

where $E_{\mathbf{P}}(q|\mathbf{S})$ is the voxel-wise Gibbs energy for a gray level q assigned to $\mathbf{P}$ and the current fixed gray levels in all neighbors of $\mathbf{P}$ in the characteristic neighborhoods $\mathbf{n}_\nu$; $\nu \in \mathbf{N}$:

$$E_{\mathbf{P}}(q|\mathbf{S}) = V_{\text{vox,nd}}(q) + \sum_{\nu \in \mathbf{N}} \sum_{(\xi,\eta,\gamma) \in \mathbf{n}_\nu} (V_{\nu,\text{nd}}(g_{x-\xi,y-\eta,z-\gamma}, q) + V_{\nu,\text{nd}}(q, g_{x+\xi,y+\eta,z+\gamma}))$$

The boundary evolution in each 2D section with the fixed z-coordinate terminates after the total energy $E_{\mathbf{r}}$ of the region $\mathbf{r} \subset \mathbf{R}$ inside the boundary $\mathbf{b}$ does not change:

$$E_{\mathbf{r}} = \sum_{\forall \mathbf{P}=(x,y,z) \in \mathbf{r}} E_{\mathbf{P}}(g_{x,y,z}|\mathbf{S}) \tag{6}$$

The deformable boundary $\mathbf{b}$ evolves in discrete time, $\tau = 0, 1, \ldots, T$, as follows:

1. Initialization ($\tau = 0$):
   (a) Initialize a boundary inside a nodule (e.g. automatically as in [2]).
   (b) Using voxels within and outside the initial boundary, estimate the current "nodule" and "background" LCDGs $\mathbf{p}_{\text{vox,nd}}$ and $\mathbf{p}_{\text{vox,bg}}$.
2. Evolution ($\tau \leftarrow \tau + 1$):
   (a) Calculate the total energy of Eq. (6) within the current boundary $\mathbf{b}_\tau$.
   (b) For each control point $\mathbf{P}_k$ on the current boundary, indicate the exterior ($-$) and interior ($+$) nearest neighbors with respect to the boundary.
   (c) For each ($+$)-point, calculate the total energy of Eq. (1) for each new candidate for the current control point.
   (d) Select the minimum-energy new candidate.
   (e) Calculate the total energy of Eq. (6) within the boundary that could have appeared if the current control point has been moved to the selected candidate position.
   (f) If the total energy increases, accept this new position of the current control point, otherwise for each ($-$)-point, calculate the total energy of Eq. (1) for each new candidate for the current control point.
   (g) Select the minimum-energy new candidate.
   (h) Calculate the total energy of Eq. (6) within the boundary that could have appeared if the current control point has been moved to the selected candidate position.
   (i) If the total energy increases, accept this new position of the current control point.
   (j) Otherwise do not move the current control point because it is already located on the edge of the desired nodule.
   (k) Mark each voxel visited by the deformable boundary.
   (l) If the current control point moves to the voxel visited earlier, then find the edge formed by the already visited voxels and use the edge points as the new control points of the deformable boundary.
   (m) If the new control points appear, interpolate the whole boundary using cubic splines and then smooth its control points with a low pass filter.
   (n) If the total energy within the boundary does not change, terminate the process; otherwise return to Step 2b.

6 Experimental Results and Conclusions

The proposed segmentation algorithm was tested on a database of clinical multislice 3D chest LDCT scans of 29 patients with $0.7 \times 0.7 \times 2.5$ mm^3 voxels that contains 350 nodules, in particular, 150 solid nodules of larger than 5 mm in diameter, 40 small solid nodules of less than 5 mm diameter, 10 cavity nodules, 61 nodules attached to the pleural surface, and 89 largely non-spherical nodules. The diameters of the nodules range from 3mm to 30mm.

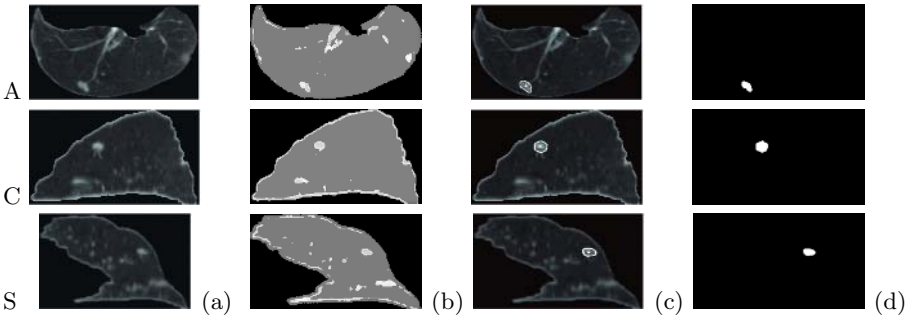


Fig. 3. 3D segmentation of pleural attached nodules; results are projected onto 2D axial (A), coronal (C), and saggital (S) planes for visualization: 2D profile of the original nodule (a), pixel-wise Gibbs energies (b) for $\nu \leq 11$, our segmentation (c), and (d) the radiologist's segmentation.

Figure 3 illustrates results of segmenting pleural attached nodules shown by axial, sagittal, and coronal cross sections. The pixel-wise Gibbs energies in each cross section are higher for the nodules than for any other lung voxels including the attached artery. Therefore, our approach separates accurately the pulmonary nodules from any part of the attached artery. The evolution terminates after 50 iterations because the changes in the total energy become close to zero. The error of our segmentation with respect to the radiologist "ground truth" is 1.86%.

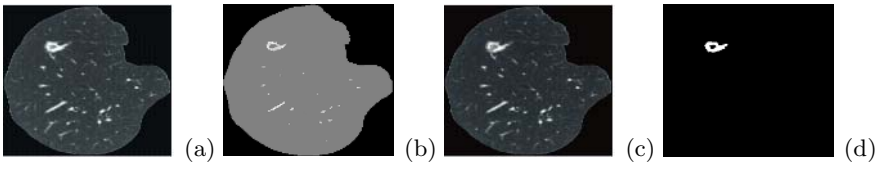


Fig. 4. 2D segmentation of cavity nodules: (a) 2D profile of the original nodule, (b) pixel-wise Gibbs energies for $\nu \leq 11$, (c) our segmentation, and (d) the radiologist's segmentation

The main advantage of our approach over the existing algorithms is in the more accurate segmentation of thin cavity nodules, i.e. the nodules that appear

only in a single slice. Experimental results in Fig. 4 show that the error of our segmentation with respect to the radiologist is 2.1%. It is worthy to note that all the existing approaches fail to segment this cavity nodule because it is totally inhomogeneous. Figure 5 presents more segmentation results obtained by our algorithm. In total, our segmentation of the 350 nodules has an error range of 0.4% – 2.35% with a mean error of 0.96%, and a standard error deviation of 1.1%.

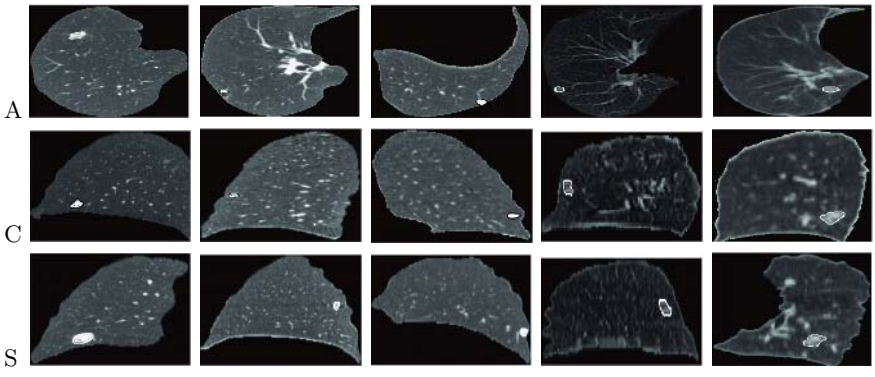


Fig. 5. Our segmentation for five more patients

We introduced a new method to accurately segment small 2D and large 3D pulmonary nodules on LDCT chest images. In our approach, the evolution of a deformable model is controlled with probability models of visual appearance of pulmonary nodules. The prior MGRF model is identified from a given training set of nodules. The current appearance model adapts the control to each bi-modal image of lung regions to be segmented. Both the models are learned using simple analytical and numerical techniques. Experiments with real LDCT chest images confirm high accuracy of our segmentation with respect to the radiologist's ground truth. Our segmentation outperforms other existing approaches for all types of nodules, in particular, for cavity nodules where other existing approaches fail.

Acknowledgement. This work has been supported by the Kentucky Lung Cancer Research Program.

References

1. A. A. Farag, A. El-Baz, and G. Gimelfarb, "Quantitative Nodule Detection in Low Dose Chest CT Scans: New Template Modeling and Evaluation for CAD System Design," *Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Palm Springs, California, USA, October 26-29, 2005, pp. 720-728.
2. A. A. Farag, A. El-Baz, and G. Gimel'farb, "Precise Segmentation of Multi-modal Images," *IEEE Transactions on Image Processing*, vol. 15, no. 4, pp. 952-968, April 2006.

3. K. Fukunaga, *Statistical Pattern Recognition*. San Diego: Academic, 1990.
4. K. Okada, D. Comaniciu, and A. Krishnan, "Robust Anisotropic Gaussian Fitting for Volumetric Characterization of Pulmonary Nodules in Multislice CT," *IEEE Trans. on Medical Imaging*, vol. 24, no. 3, pp.409–423, March 2005.
5. B. Zhao, D. Yankelevitz, A. Reeves, and C. Henschke, "Two-dimensional multicriterion segmentation of pulmonary nodules on helical CT images," *IEEE Trans. on Medical Imaging*, vol. 22, pp. 1259–1274, 2003.
6. W. J. Kostis, A. P. Reeves, D. F. Yankelevitz, and C. I. Henschke, "Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images," *IEEE Trans. on Medical Imaging*, vol. 22, pp. 1259–1274, 2003
7. M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *International Journal of Computer Vision*, vol. 1, pp. 321–331, 1987.

Intensity-Based Volumetric Registration of Contrast-Enhanced MR Breast Images

Yin Sun[1], Chye Hwang Yan[1], Sim-Heng Ong[1,2],
Ek Tsoon Tan[1], and Shih-Chang Wang[3]

[1] Department of Electrical and Computer Engineering,
National University of Singapore
[2] Division of Bioengineering, National University of Singapore
[3] Department of Diagnostic Radiology, National University of Singapore

Abstract. In this paper, we propose a fast intensity-based registration algorithm for the analysis of contrast-enhanced breast MR images. Motion between pre-contrast and post-contrast images has been modeled by a combination of rigid transformation and free-form deformation. By modeling the conditional probability function to be Gaussian and considering the normalized mutual information (NMI) criterion, we create a pair of auxiliary images to speed up the registration process. The auxiliary images are registered to the actual images by optimizing the simple sum of squared difference (SSD) criterion. The overall registration is achieved by linearly combining the deformation observed in the auxiliary images. One well-known problem of non-rigid registration of contrast enhanced images is the contraction of enhanced lesion volume. We address this problem by rejecting the intensity outliers from registration. Results have shown that our method could achieve accurate registration of the data while successfully prevent the contraction of the contrast enhanced lesion volume.

1 Introduction

Contrast-enhanced magnetic resonance mammography (CEMRM) uses magnetic resonance imaging (MRI) to obtain 3-D tomography of the human breast. An intravenously-injected, paramagnetic contrast agent (Gd-DTPA) enhances vascular structures, including hypervascular lesions such as breast cancers. When several images are obtained in a time sequence, malignancy may be distinguished by the enhancement curve of each voxel. However such analysis cannot always be directly applied since patient motion due to breathing and discomfiture is present. The breast is also soft and deformable and will not move in a uniform fashion. To analyze these images, we need to register them. Breast image registration has been studied by several research groups. Some previous attempts at registering CEMRM images used mutual information (MI) or normalized mutual information (NMI) as a similarity measure [1, 2, 3]. MI and NMI measures the statistical dependency between pair of images and therefore they are insensitive to intensity changes. To model the motion, it has already been shown

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 671–678, 2006.

that free-form deformation is a viable tool to effectively reduce the motion artifacts that exist in CEMRM images as shown by Rueckert *et al.* in [3]. They used a combination of global affine and local free-form transformation to model the image deformation. A known problem of CEMRM image registration is the artificial reduction of the volume of contrast enhanced lesions due to intensity changes [2]. This is definitely undesirable since breast tissue is known to be incompressible.

The determination of MI and NMI is a very computation-intensive task. It requires the formation of the joint histogram of corresponding voxel pairs. The optimization of the transformation parameters often requires computation of the gradient of the MI or NMI based cost function with respect to the transformation parameters. With appropriate interpolation of the histogram, an analytic expression could be computed for MI derivatives. Maes *et al.* used partial volume interpolation on the histogram and derived analytic derivatives of MI to allow exact computation of the gradient [4]. Thévenaz *et al.* proposed to use Parzen windowing to form the histogram and derived an analytic form for the MI gradient [5]. Another way of computing the gradient is through numerical approximation. Wells *et al.* used stochastic approximation for the MI gradient [6]. Rohlfing *et al.* used finite-difference approximation to find the gradients of NMI [7]. Stochastic approaches only make use of limited samples from the data, not all information from the data has been utilized. Other approaches involving all samples are very computationally intensive. During optimization of the transformation, the gradient of the cost function is often required to be computed many times. Therefore, it is desirable to design fast methods to approximate the gradients.

In this paper, we propose a simplified way to register CEMRM images. The sum of squared difference (SSD) is a robust and fast when intensity is assumed constant [8, 9]. If we can correct for the intensity changes in CEMRM, SSD can be used as the similarity measure. Motivated by this observation, we show that if the conditional probability function is parameterized as Gaussian, NMI can be reduced to a combination of SSD terms. Thus, we divide the overall registration problem into two subproblems. Each of the subproblem involves the registration of an auxiliary image to the original image using SSD. With the final solution obtained from the solutions to the subproblems. To further enhance the robustness of our method, we identify intensity outliers and exclude them from non-rigid registration. This is simply done by thresholding the intensity difference when computing SSD, and the effect of outliers is then reduced significantly.

2 Method

We perform registration in two stages, i.e., rigid registration followed by nonrigid registration. Spline basis functions are used in nonrigid registration. In each stage of registration, the problem is divided into the subproblems of registering auxiliary images to the original images. Overall registration is achieved by combining them. In the nonrigid registration phase, we reject the outliers in order to prevent erroneous local deformation.

2.1 Problem Formulation

We denote the pre-contrast image to be U and the post-contrast image to be V. Defining some similarity measure $S(\cdot)$ that is either minimized or maximized when the images are aligned, we could formulate the overall registration as

$$T = \arg\max_{T}[S(u(\mathbf{p}), v(T(\mathbf{p})))] \tag{1}$$

$T(\cdot)$ is some geometrical transformation that models the motion between images, and $\mathbf{p}$ the coordinates of a voxel in the reference image. The optimal transformations will be represented by a T that gives a similarity measure S at its optimal point. S could be any similarity measure that gauges the registration quality. The simplest will be the SSD criterion, which is attractive because of it low complexity.

$$\mathrm{SSD}(u, v) = \frac{1}{N}\sum_{\mathbf{p}}[u(\mathbf{p}) - v(T(\mathbf{p}))]^2 \tag{2}$$

where N is the total number of voxels. Another option which does not require intensity conservation is NMI.

$$\mathrm{NMI}(u, v) = \frac{H(u(\mathbf{p})) + H(v(T(\mathbf{p})))}{H(u(\mathbf{p}), v(T(\mathbf{p})))} \tag{3}$$

where $H(\cdot)$ denotes the marginal entropy and $H(\cdot,\cdot)$ the joint entropy. In the following subsection, we describe how NMI may be simplified to SSD so as to achieve fast registration.

2.2 Dividing Registration Task into Subproblems

It has been shown that the joint probability density function (joint PDF) of the image voxel pairs can be modeled as a Gaussian mixture [10]. It is difficult to estimate the Gaussian mixture due to the large number of parameters need to be determined. Instead of modeling the joint probability function, we model the conditional probability density function (conditional PDF) between image voxel pairs.

The conditional PDF is given using the standard Gaussian density function, $p(u|v) = \frac{1}{\sqrt{2\pi}\sigma_v}\exp\{-\frac{(u-m_v)^2}{2\sigma_v^2}\}$, where a pair of volumes are denoted by $u(\mathbf{p})$ and $v(T(\mathbf{p}))$, we have omitted the coordinate $\mathbf{p}$ for simplicity. m_v and σ_v^2 denote the conditional mean and variance respectively. By considering the conditional entropy of $u(\mathbf{p})$ and $v(T(\mathbf{p}))$, it follows that

$$H(u|v) = \frac{1}{2}\log(2\pi) + \sum_{v} p(v)\log(\sigma_v) + \mathrm{SSD}(\hat{u}, \hat{v}) \tag{4}$$

where $\mathrm{SSD}(\cdot)$ denotes the sum of squared difference as defined in (2) and we have introduced a pair of normalized signals $\hat{u}(\mathbf{p}) = \frac{u(\mathbf{p})}{\sqrt{2}\sigma_v}$ and $\hat{v}(\mathbf{p}) = \frac{m_v}{\sqrt{2}\sigma_v}$. In (4), the first term is constant. If we can assume the second term involving σ_v to be

constant, we can greatly simplify the computation of the derivative. σ_v turns out to be approximately constant during nonrigid registration because the extent of local motion is extremely small. Similarly, we can derive another expression for $H(v(T(\mathbf{p}))|u(\mathbf{p}))$ and the corresponding pair of approximated signals $\tilde{u}(\mathbf{p})$ and $\tilde{v}(T(\mathbf{p}))$. It follows that an approximation of the NMI gradient with respect to a transformation variable ϕ is given by

$$\frac{\partial \mathrm{NMI}(u,v)}{\partial \phi} = \frac{1}{H^2(u,v)}[-\mathrm{MI}(u,v)\frac{\partial}{\partial \phi}\mathrm{SSD}(\tilde{u},\tilde{v}) - H(u,v)\frac{\partial}{\partial \phi}\mathrm{SSD}(\hat{u},\hat{v})] \quad (5)$$

where $\mathrm{MI}(u,v) = H(u) + H(v) - H(u,v)$ denotes the mutual information of the two images. $\hat{u}(\mathbf{p})$ and $\hat{v}(T(\mathbf{p}))$ can be interpreted as the normalized images with respect to the pre-contrast image. The contrast enhanced structures in $v(\mathbf{p})$ would be replaced by non-enhanced structures expected to be found in the pre-contrast image. On the other hand, $\tilde{u}(\mathbf{p})$ and $\tilde{v}(T(\mathbf{p}))$ could be regarded as normalized images with respect to the post-contrast image. These images are regarded as auxiliary images which will aid us in registration. From (5), we can see that the two SSD terms involved are weighted by MI and JE *(joint entropy)* respectively. At the initial stage of registration, JE will be large and MI will be small because the two images are not well aligned. Therefore, the value of the gradient will be dominated by the second SSD term. With better alignment, we expect JE to decrease and MI to increase. The contribution of the two SSD terms will be weighted accordingly. When computing SSD, we consider the effect of intensity outliers which are not removed by the normalization. SSD is computed as,

$$\mathrm{SSD}(u,v) = \frac{1}{N}\sum_{\mathbf{p}} \min[|u(\mathbf{p}) - v(T(\mathbf{p}))|, \xi]^2 \quad (6)$$

where N is the number of voxels and ξ is a threshold used to suppress the effect of outliers. Experimentally, we found that a threshold value of 100 gives satisfactory results.

It is interesting to note that each of the SSD terms is in fact equivalent to the computation of correlation ratio (CR), introduced by Roche *et al.* [11]. The existence of equivalence provides new insights to the different similarity measures used. Although our starting point is NMI, the derived expression of NMI (5) combines MI, JE and CR in an elegant manner.

3 Experiments

3.1 Data

We used a total of 17 normal patient datasets and 10 lesion datasets. Image acquisition was done using a GE Sigma 1.5 Tesla coil MRI scanner with 3-D fast-spoiled gradient echo and no spectral fat suppression (TR = 25.6ms, TE=3ms, fractional echo, Flip angle = 30^o, FOV = 32 to 40cm). The contrast agent

used was MagneVist Gd-DTPA of concentration 0.2mmol/kg. A typical dataset has 5 scans (256×256×24 voxels) of voxel size 1.05mm×1.05mm×5.45mm. Slice direction used is axial. Variations to this protocol are mainly in the number of slices, which can vary from 16 to 56 depending on the volume size to be acquired. The contrast agent is injected after the first scan, and post-contrast scans will follow in the next 5 to 20 minutes. Each 3-D scan requires 30-60 seconds of acquisition time, depending on the number of slices.

3.2 Implementation

We have implemented the registration algorithm based on the Insight Toolkit [12]. A rectangular region of interest (ROI) around each breast is manually selected and ROI registration is performed. Rigid registration is implemented using gradient descent optimization with a single resolution. Optimization is terminated when the change in cost function is smaller than a predefined threshold value. We have found that a threshold between 10^{-2} to 10^{-4} could achieve good registration. Nonrigid registration also employs gradient descent technique with four different image resolution used. The terminating condition for nonrigid registration will also depend on the value of the cost function. If the change in value of the cost function is smaller than a threshold, we stop the optimization process. Currently, the running time is about 10 to 15 minutes. However, we expect this time to further decrease with code optimization.

3.3 Registration Quality

It is difficult to provide a numerical test of registration quality. We have identified sum of squared difference (SSD) (2), normalized mutual information (NMI) (3) and normalized correlation coefficient ($\mathrm{NCC}(u,v) = \frac{\sum_{\mathbf{p}}(u(\mathbf{p})\cdot v(T(\mathbf{p})))}{\sqrt{\sum_{\mathbf{p}} u(\mathbf{p})^2 \cdot \sum_{\mathbf{p}} v(T(\mathbf{p}))^2}}$) as similarity measures to gauge how well the two images match one another. Since our NMSSD method does not optimize these three similarity measures directly, the value of these measures could give us an indication of the objective registration quality. Across the board, rigid registration is able to gain in terms of similarity measures compared to no registration for both NMSSD method and SSD method. Similarly, nonrigid registration further improves registration for both methods. As shown in Table 1 and Table 2, the NMSSD method generally offers greater gain over traditional SSD method for all the similarity measures. We have separated the lesion breasts from the normal breasts in our tests. Results have shown that NMSSD performs consistently better than SSD for both normal cases and lesion cases.

The gain of NMSSD over SSD could be attributed to the use of auxiliary images, which accounts for the intensity changes induced by the contrast agent. This will reduce the chance of the optimization being trapped in a local minima. The chances of having erroneous registration is reduced too since we corrected the intensity difference.

Table 1. Registration results of NMSSD method. The average value and the variance of SSD, NMI and NCC is presented separately for a) before registration; b) after rigid registration; c) after nonrigid registration.

	Normal Cases			Lesion Cases		
	Before	Rigid	Nonrigid	Before	Rigid	Nonrigid
SSD (Average)	51.896	39.496	33.508	47.681	39.713	34.177
(Variance)	62.304	54.875	56.173	31.453	28.545	26.459
NMI (Average)	1.2699	1.2849	1.2934	1.2459	1.2625	1.2698
(Variance)	0.0797	0.0751	0.0762	0.0450	0.0426	0.0425
NCC (Average)	0.9803	0.9832	0.9841	0.9916	0.9933	0.9941
(Variance)	0.0413	0.0379	0.0380	0.0057	0.0047	0.0045

Table 2. Registration results of traditional SSD method. The average value and the variance of SSD, NMI and NCC is presented separately for a) before registration; b) after rigid registration; c) after nonrigid registration.

	Normal Cases			Lesion Cases		
	Before	Rigid	Nonrigid	Before	Rigid	Nonrigid
SSD (Average)	51.896	41.838	31.355	47.681	42.197	34.867
(Variance)	62.304	55.991	44.844	31.453	30.636	27.431
NMI (Average)	1.2699	1.2798	1.2917	1.2459	1.2577	1.2684
(Variance)	0.0797	0.0759	0.0794	0.0450	0.0437	0.0427
NCC (Average)	0.9803	0.9825	0.9841	0.9916	0.9929	0.9940
(Variance)	0.0413	0.0387	0.0373	0.0057	0.0051	0.0047

3.4 Effect on Lesion Volume Reduction

To calculate the volume of the lesion, we create a binary mask of the breast lesion after rigid registration. The lesion is identified using the 3TP method [13]. The volume of the lesion is determined from the total number of voxels in the mask that are labelled as lesion. To see whether there are any changes in the lesion volume, we apply the final transformation obtained from the optimization process to the mask, and compute the lesion volume in terms of number of lesion voxels after transformation. The change in volume could then be calculated as

$$\Delta V = \frac{V_{after} - V_{before}}{V_{before}} \times 100\% \tag{7}$$

where ΔV denotes the change in volume, V_{after} and V_{before} are the number of lesion voxels before and after nonrigid registration. We measure the volume change for the case where outlier rejection is performed and compare it with the case where outliers are included. The effect of lesion reduction is only evident if the size of the lesion is not too small. In our experiments, we pick from the lesion datasets a few examples that contains big lesions and tested the effect of volume reduction.

Table 3. Effect of lesion volume reduction in terms of number of lesion voxels before and after nonrigid registration

	Outliers Included			Outliers Rejected		
Case ID	Before	After	Reduction	Before	After	Reduction
1	1816	1464	-19.4%	1816	1799	-0.936%
2	383	348	-9.14%	383	371	-3.13%
3	1173	1099	-6.31%	1173	1140	-2.81%
4	2561	1754	-31.5%	2561	2223	-4.80%
5	276	257	-6.88%	276	263	-4.71%

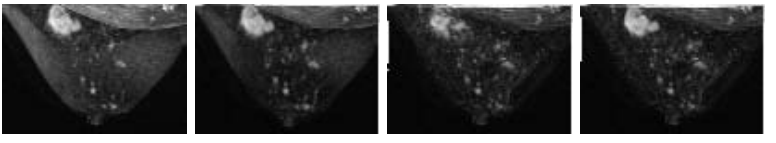


Fig. 1. Lesion volume test Case 1. From left to right: Maximum intensity projection (MIP) of the subtraction image without registration; MIP of subtraction image after rigid registration; MIP of subtraction image after nonrigid registration without lesion segmentation, the volume shrank by 19.38%; MIP of subtraction image after nonrigid registration with lesion segmentation, volume shrinkage is reduced to 0.936%.

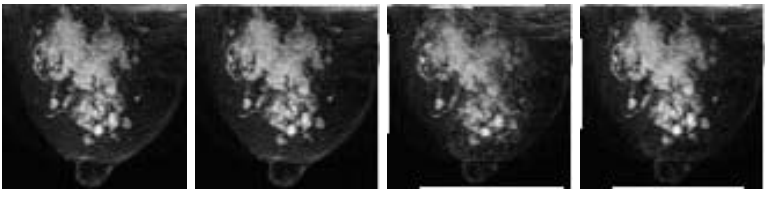


Fig. 2. Lesion volume test Case 4. From left to right: Maximum intensity projection (MIP) of the subtraction image without registration; MIP of subtraction image after rigid registration; MIP of subtraction image after nonrigid registration without lesion segmentation, the volume shrank by 31.51%; MIP of subtraction image after nonrigid registration with lesion segmentation, volume shrinkage is reduced to 4.802%.

The results are presented in Table 3. By excluding the lesions from optimization, it is possible to preserve the lesion volume. There is no significant volume reduction observed. However, due to the registration of surrounding regions, slight contraction of the volume in the order of less than 5% of the original volume is observed. The contraction is extremely small, and hence it can be considered negligible.

4 Conclusion

We have developed a nonrigid registration algorithm for CEMRM images in this paper. By the use of auxiliary images constructed from conditional probability

distribution of image pairs, we showed that the problem could be divided into subproblems of registering the auxiliary images to original images using SSD. Our algorithm is able to take into account of the non-uniform intensity changes induced by the administration of contrast agents, thus reducing the chances of misregistration. Experiments have confirmed the effectiveness of our algorithm. Our algorithm could consistently outperform traditional sum of squared difference. By excluding outliers from nonrigid registration, we prevent artificial lesion volume reduction. Experiments have shown that this method is simple yet effective. We can restrict the change of lesion volume to within 5% of its original volume.

References

1. Rohlfing, T., Jr., C.R.M., Bluemke, D.A., Jacobs, M.A.: Volume-preserving nonrigid registration of mr breast images using free-form deformation with an incompressibility constraint. IEEE Trans. Med. Imaging **22**(6) (2003) 730–741
2. Tanner, C., Schnabel, J.A., Chung, D., Clarkson, M.J., Rueckert, D., Hill, D.L.G., Hawkes, D.J.: Volume and shape preservation of enhancing lesions when applying non-rigid registration to a time series of contrast enhancing mr breast images. In: MICCAI. (2000) 327–337
3. 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) (1999)
4. Maes, F., Vandermeulen, D., Suetens, P.: Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information. Medical Image Analysis **3**(4) (1999) 373–386
5. Thévenaz, P., Unser, M.: Optimization of mutual information for multiresolution image registration. Volume 9. (2000) 2083–2099
6. Wells, W., Viola, P., Atsumi, H., Nakajima, S., Kikinis, R.: Multi-modal volume registration by maximization of mutual information. Medical Image Analysis **1**(1) (1996) 35–51
7. Rohlfing, T., Jr., C.R.M.: Nonrigid image registration in shared-memory multiprocessor environments with application to brains, breasts, and bees. IEEE Transactions on Information Technology in Biomedicine **7**(1) (2003) 16–25
8. Kybic, J., Unser, M.: Fast parametric elastic image registration. IEEE Transactions on Image Processing **12**(11) (2003) 1427–1442
9. Clarenz, U., Droske, M., Rumpf, M.: Towards fast non-rigid registration. Comtemporary mathematics, Special issue on inverse problems and image analysis (2002)
10. Leventon, M.E., Grimson, W.E.L.: Multi-modal volume registration using joint intensity distributions. In: MICCAI. (1998) 1057–1066
11. Roche, A., Malandain, G., Pennec, X., Ayache, N.: The correlation ratio as a new similarity measure for multimodal image registration. In: MICCAI. (1998) 1115–1124
12. InsightToolkit: (http://www.itk.org)
13. Kelcz, F., Furman-Haran, E., Grobgeld, D., Degani, H.: Clinical testing of high-spatial-resolution parametric contrast-enhanced mr imaging of the breast. American Journal of Radiology (**179**)

Semi-parametric Analysis of Dynamic Contrast-Enhanced MRI Using Bayesian P-Splines

Volker J. Schmid[1], Brandon Whitcher[2], and Guang-Zhong Yang[1]

[1] Institute for Biomedical Engineering, Imperial College, South Kensington, London SW7 2AZ, United Kingdom
{v.schmid, g.z.yang}@imperial.ac.uk
[2] Translational Medicine & Genetics, GlaxoSmithKline, Greenford Road, Greenford UB6 0HE, Middlesex, United Kingdom
brandon.j.whitcher@gsk.com

Abstract. Current approaches to quantitative analysis of DCE-MRI with non-linear models involve the convolution of an arterial input function (AIF) with the contrast agent concentration at a voxel or regional level. Full quantification provides meaningful biological parameters but is complicated by the issues related to convergence, (de-)convolution of the AIF, and goodness of fit. To overcome these problems, this paper presents a penalized spline smoothing approach to model the data in a semi-parametric way. With this method, the AIF is convolved with a set of B-splines to produce the design matrix, and modeling of the resulting deconvolved biological parameters is obtained in a way that is similar to the parametric models. Further kinetic parameters are obtained by fitting a non-linear model to the estimated response function and detailed validation of the method, both with simulated and *in vivo* data is provided.

1 Introduction

Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) has become an important source of information to aid cancer diagnosis. After the administration of a contrast agent, such as *Gadolinium diethylenetriaminepentaacetic acid* (Gd-DTPA), a dynamic sequence of MR scans is acquired, typically using T_1-weighted sequences for assessing the reduction in T_1 relaxation time caused by the contrast agent. The contrast concentration may be estimated from the observed signal intensity using proton density-weighted images after calibration [1]. Quantitative methods based on pharmacokinetic (PK) models are then used to fit a non-linear function to the estimated contrast agent concentration. These models are usually derived from the solution to a system of linear differential equations [2].

A two-compartment model describing flow between plasma (or vascular space) and extra-vascular extra-cellular space (EES) is commonly used. Mathematically, this can be described as

$$C_t(t) = C_p(t) \otimes K^{\text{trans}} \exp(-k_{\text{ep}} t). \tag{1}$$

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 679–686, 2006.

where C_t is the concentration of contrast agent in the tissue at time t, C_p the arterial input function (AIF) and the kinetic parameters of interest are K^{trans} and k_{ep}. If the AIF is given in parametric form, the convolution in (1) may be written in an analytical form and the contrast agent concentration can be estimated by using standard non-linear regression methods [3].

Recently, non-parametric or model-free techniques for DCE-MRI have received extensive attention. These include the use of neural networks [4] for tissue classification, and semi-parametric methods for kinetic modeling of dynamic PET imaging [5]. Jerosch-Herold *et al.* [6] have proposed a model-free approach for myocardial perfusion quantification, based on a B-spline polynomial representation of the response function. The formulation of this problem, however, is ill-conditioned and a first-order difference penalty was imposed.

Penalty splines, or P-splines, represent another popular approach for semi-parametric modeling [7].The underlying function is approximated by a linear combination of a relatively large number of B-spline basis functions, where smoothness is ensured by a penalty based on k-th order differences of the parameter vector. For selecting the penalty weight (or smoothing parameter) λ cross validation can be used. Bayesian versions of P-splines estimate λ jointly with the model and allow for varying weights; *i.e.* adaptive smoothing [8].

In practice, an adaptive approach to modeling the contrast agent concentration in DCE-MRI is preferred since the observed time series vary rapidly in the first two minutes after bolus injection of the contrast agent. The purpose of this paper is to apply Bayesian P-spline regression to the estimated contrast agent concentration, proposing an ad-hoc estimator for plasma flow, derived directly from the response function. Further kinetic parameters are obtained by fitting a non-linear model to the estimated response function. Detailed validation of the method both with simulated data and clinical cases of breast tumor is provided.

2 Theory and Methods

2.1 Theoretical Background

The time series of measured contrast agent concentration $C_t(t)$ is the convolution of the arterial input C_p and the response function f, such that

$$C_t(t) = C_p(t) \otimes f(t) = \int_0^t C_p(t-u) f(u)\, du. \tag{2}$$

The concentration is measured at discrete time points $\tau_1, \ldots, \tau_n$. The AIF may be measured (or estimated) at a different resolution $t_1, \ldots, t_T$, or a standardized function as *e.g.* proposed by Tofts and Kermode [9] may be used. The discretized form of (2) is given by

$$C_t(\tau_i) = \sum_{j=1}^{T} C_p(\tau_i - t_j) f(t_j) \Delta t = \sum_{j=1}^{T} A_{ij} f(t_j), \tag{3}$$

where Δt represents the sampling interval of the AIF. The $T \times n$ matrix $\boldsymbol{A}$ may be interpreted as a convolution operator and is defined via

$$A_{ij} = \begin{cases} C_p(\tau_{n_i - j+1})\Delta\, t & \text{if } t_i < \tau_j; \\ 0 & \text{else,} \end{cases} \tag{4}$$

where n_i is the maximum index j for which $t_j \leq \tau_i$ holds. To improve numerical stability, a B-spline representation for the response function is used in this study such that

$$f(t) = \sum_{j=1}^{p} \beta_j B_{jt}, \tag{5}$$

where $\boldsymbol{B}$ is the $n \times p$ design matrix of kth order B-splines with knots $s_1, \ldots, s_{p+k}$. More knots are placed in the first two minutes (about every 3-5 seconds) and fewer afterwards (about every 20-30 seconds) with a total of approximately 40-50 knots. In vector notation $\boldsymbol{f} = (f(t_1), \ldots, f(t_T))'$ and (5) may be expressed via $\boldsymbol{f} = \boldsymbol{B\beta}$. When we put this into (3), we get

$$\boldsymbol{C_t} = \boldsymbol{Af} = \boldsymbol{AB\beta} = \boldsymbol{D\beta}, \tag{6}$$

with $\boldsymbol{D} = \boldsymbol{AB}$ a $T \times p$ design matrix, which is the (discrete) convolution of the AIF with the B-spline polynomials.

2.2 Bayesian Model

A fully Bayesian hierarchical model is used to fit the proposed semi-parametric model. Following the notation in the previous section, the corresponding *data model* in each voxel becomes

$$C_t(t) = \boldsymbol{D}_t\boldsymbol{\beta} + \epsilon_t \quad \text{for each } t, \tag{7}$$

where $\boldsymbol{D}_t$ denotes the tth row of $\boldsymbol{D}$, with $\boldsymbol{D} = \boldsymbol{AB}$.

Although the system of equations is linear in $\boldsymbol{\beta}$, (7) may still be numerically unstable. A stochastic constraint on $\boldsymbol{\beta}$ is introduced in the *process model* using a second-order random walk prior, to aquire a reasonably smooth representation of the response function. That is, starting from the third coefficient, each coefficient is *a priori* assumed to be the linear continuation of the last two coefficients

$$\beta_i = 2\beta_{i-1} - \beta_{i-2} + e_i \qquad \text{for } i = 3, \ldots, p. \tag{8}$$

The inverse variance (precision) δ_i of the error term e_i in (8) determines the smoothness of the spline. A non-informative hyperprior is proposed, $e_t \sim \mathrm{N}(0, 1/\delta_t)$, $\delta_t \sim \mathrm{Ga}(10^{-5}, 10^{-5})$, where Ga denotes the Gamma distribution. As the response function exhibits a sharp increase and a sharp decrease right at the beginning of the dynamic series, the smoothness parameter is estimated adaptively, *i.e.*, the precision δ_t is allowed to vary over time.

For the observation error ϵ we assume *a priori* a Normal distribution with unknown variance, $\epsilon_t \sim \mathrm{N}(0, 1/\kappa)$, $\kappa \sim \mathrm{Ga}(1, 0.05)$. Bayes' theorem is applied to calculate the posterior distribution of the parameter vector. We use a Markov chain Monte Carlo (MCMC) algorithm to obtain samples from the posterior distribution [10].

2.3 Obtaining Kinetic Parameters

In order to obtain kinetic parameters, the mean plasma flow F_p is estimated via the maximum of the response function. This is equivalent to the suggestions of Jerosch-Herold *et al.* [6], who use the maximum of the response function to estimate the maximum blood flow in myocardial perfusion.

To obtain additional kinetic parameters, a non-linear model is fitted to the response function. St. Lawrence and Lee [11] have proposed a step function in their adiabatic approximation to the tissue homogeneity model. In practice, a step function is highly unlikely to be observed especially in the smoothing framework presented here, so a linear approximation is used instead given by

$$H(t) = \begin{cases} 1 - t\,(1-E)/\tau & \text{for } 0 \le t < \tau; \\ E\exp[-(t-\tau)EF_p/v_e] & \text{for } t \ge \tau, \end{cases} \tag{9}$$

and $f(t) = F_p \cdot H(t)$. Here $\tau = 2T_C$, the transit time through the capillary, E is the extraction fraction of the tracer that is extracted into the EES during a single capillary transit and v_e is the volume fraction of EES.

It is worth noting that the Bayesian methodology in Section 2.2 provides not just one response function, but a distribution on the space of response functions. By fitting (9) to each response function in the sample we obtain a distribution of the estimated parameters. The median of the posterior distributions is used to estimate the parameters, the estimation error can be estimated via the standard errors of the samples and intervals can be constructed from quantiles of the posterior distributions.

3 Results

3.1 Validation on Simulated Data

To assess the underlying accuracy of the method, the proposed method is applied to simulated data provided by David Buckley [12] with known kinetic parameters. The data was simulated using MMID4, part of a software made available by the National Simulation Resource, Dept. of Bioengineering, University of Washington (`http://www.nsr.bioeng.washington.edu`). We use the first series from [12], which was designed to be representative of data acquired from a breast tumor.

As a reference parametric model, an extended version of the modified Kety model is used

$$C_t(t) = v_p C_p(t) + K^{\text{trans}} \int_0^t C_p(u)\exp[-k_{\text{ep}}(t-u)]\,du, \tag{10}$$

which adds a vascular term v_p (plasma volume fraction) to the standard model, see (1). Note, that $K^{\text{trans}} = EF_p$, $k_{\text{ep}} = K^{\text{trans}}/v_e$ and $v_p = T_c/F_p$.

As one would expect, the goodness of fit is much better with the semi-parametric technique than with the parametric model. The mean of the residual

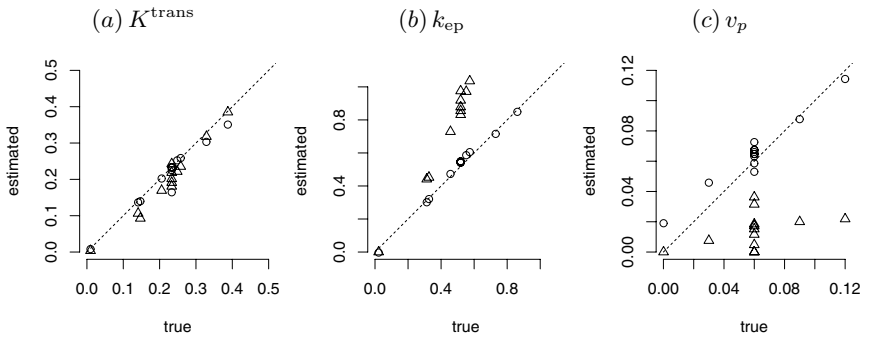


Fig. 1. Scatter plot of parameter estimates vs. true values. Estimates of the parameters (a) K^{trans}, (b) k_{ep} and (c) v_p. Circles represent the results using the semi-parametric method, triangles the results using the parametric technique.

sums of squares for the semi-parametric method, over all 13 experiments, was $1.13 \cdot 10^{-3}$ (ranging from $1.89 \cdot 10^{-4}$ to $2.13 \cdot 10^{-3}$). This implies that the data was nearly interpolated by this model. The mean of the residual sums of squares for the parametric model was 0.661 (ranging from 0.198 to 1.520).

Fig. 1 shows the parameter estimates for both methods as a function of the true values. The parameter K^{trans} is underestimated with the parametric model (in average by 17.7%, consistent with previously reported results, [12]), whereas the semi-parametric technique is doing much better (mean deviation from true values is 6.2%). The parameter k_{ep} is strongly overestimated with the parametric method (between 38% and 100%), as v_p is underestimated (by 39% up to 100%). With the semi-parametric approach, k_{ep} is slightly overestimated only (by 1.5% to 6.2%), whereas v_p is estimated well at 11 out of the 13 experiments (with error of 2.5% to 21%), whereas two values show a less good fit (52% error and a value of 0.18 instead of 10^{-4}). Overall, the parameters obtained from the semi-parametric approach fit much better to the ground truth.

The maximum response function (MRF) can be used to estimate F_p. The values of the MRF and the estimation of F_p via non-linear fitting from the estimated response function are nearly equivalent, with a correlation of 99.94%. Fig. 2(a) shows both the non-linear fitting from the response function and the MRF plotted against the true values. With the proposed Bayesian technique, not only are the parameters estimated, but estimations of the error or interval estimators are also available. Fig. 2(b) gives 95% credibile intervals for K^{trans}. When the point estimate is relatively far away from the true value (Exps. 6, 12 and 13) the intervals are wide and cover the true value, indicating that the technique provides important information about both the accuracy and precision of its estimates.

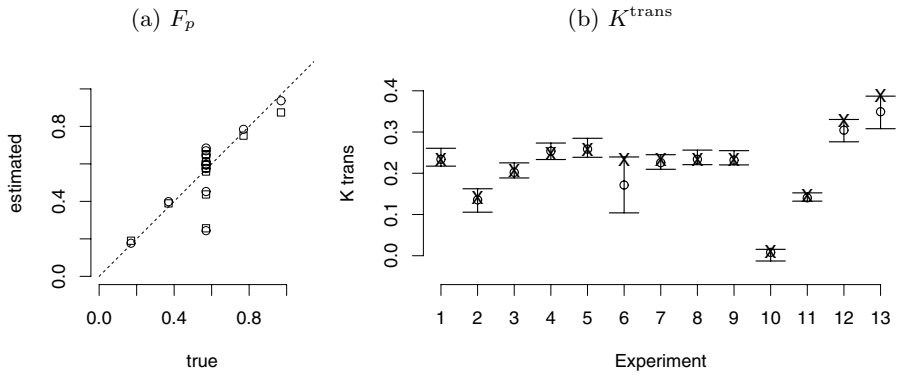


Fig. 2. (a) Scatter plot of true values against estimated values of F_p with ad-hoc method (squares) and non-linear fitting (circles). (b) Estimates of K^{trans} (circles) with 95% intervals; true values are marked as cross.

3.2 In Vivo Validation

The clinical data used in this study were provided by the Paul Strickland Scanner Centre at Mount Vernon Hospital, Northwood, UK. The data were derived from six patients with breast tumor scanned with a 1.5 T Siemens MAGNETOM Symphony scanner, TR = 11 ms and TE = 4.7 ms; 40 scans with four sequential slices were acquired in about 8 minutes. A dose of D = 0.1 mmol/kg body weight Gd-DTPA was injected at the start of the fifth acquisition using a power injector. One slice from the pre-treatment scan of each patient was used for analysis. Data from this study have been previously reported [13, 3].

Figs. 3(a)–(c) show the observed contrast concentration time series from three voxels (tumor, peri-tumoral rim, normal tissue) together with the fits from (10) and from the proposed semi-parametric approach. It is evident that in all vox-

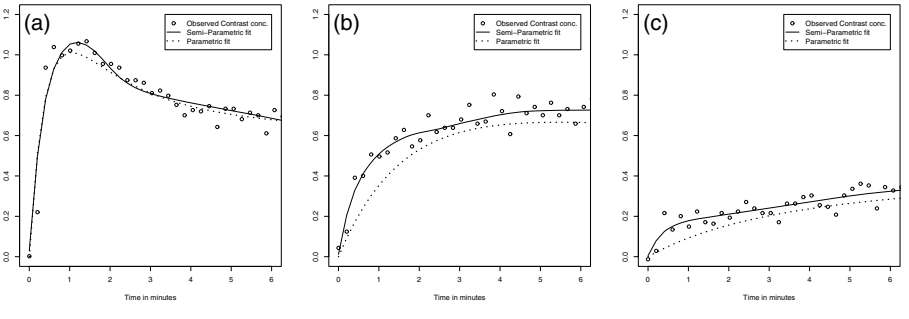


Fig. 3. Observed concentration time series (circles) with fit by parametric (dotted line) and semi-parametric method (solid line) for three voxels: (a) tumor, (b) peri-tumoral rim, (c) normal tissue

Table 1. Sum of squared error (SQE) between data and fitted function for parametric and semi-parametric method, averaged over all voxels. Last row - Mean percentage of K^{trans} parameter estimated with semi-parametric approach in respect of values estimated with parametric approach (for tumor voxels)

Patient	1	2	3	4	5	6
SQE, par.	0.0709	0.1267	0.3992	0.1747	0.3432	0.1730
SQE, semi-par.	0.0515	0.0614	0.1054	0.0443	0.0844	0.0876
perc.K^{trans}	195%	118%	564%	801%	297%	142%

els the semi-parametric method fits the data much better than the parametric approach. This is to be expected since the degrees of freedom of the model in the semi-parametric approach are much higher. The figures also indicate that the standard parametric model is not always appropriate for the observed data, where the initial upslope is difficult to fit using the parametric model.

Finally, Table 1 shows the goodness of fit for the data from all six patients and the mean percentage of K^{trans} obtained from the semi-parametric and the parametric approach. As with the simulated data, the fit is clearly better with the semi-parametric technique. Also, K^{trans} values are again higher with the semi-parametric method, showing a similar trend as the simulation study.

4 Conclusion

In this paper, we have presented a Bayesian P-spline model for quantifying contrast agent concentration time series of DCE-MRI scans. Compared with parametric PK models, the proposed semi-parametric technique provides a much better fit to the data. In particular the proposed method fits the important upslope of the time series more precisely. Parametric non-linear models are typically difficult to estimate due to the convergence issues and problems in specifying consistent starting values. Bayesian non-linear regression can overcome convergence issues, but with an increase in computation time [3]. The proposed semi-parametric method provides a reliable, practical algorithm by effectively circumventing the above problems.

In contrast to classical approaches [6], Bayesian P-splines allow the simultaneously estimation of model and smoothing parameters. Adaptive smoothing parameters can easily be obtained, and are important to model the sharp changes in the dynamic series.

The proposed method allows a direct assessment of the response function, *i.e.*, the actual flow in the tissue. Parameters of interest may be estimated by an ad-hoc technique (maximum of response function) or by fitting a non-linear model to the response function. In the latter case, smoothing via P-splines provides error reduction and deconvolution of the arterial input function. Results from simulated data show a clear improvement in parameter estimation. From an initial clinical validation, a similar trend of differences between parameters estimated with parametric and semi-parametric approach was observed. The

proposed method also provides standard errors for all parameters, and therefore the reliability of the results can be quantified.

Acknowledgements

Support for Volker Schmid was financed through a research grant from GlaxoSmithKline. We are grateful to David Buckley for providing the simulated data. The clinical data was graciously provided by Anwar Padhani and Jane Taylor.

References

1. Parker, G.J.M., Padhani, A.R.: T_1-w DCE-MRI: T_1-weighted dynamic contrast-enhanced MRI. In Tofts, P., ed.: Quantitative MRI of the Brain. Wiley, Chichester, England (2003) 341–364
2. Larsson, H.B., Tofts, P.S.: Measurement of the blood-brain barrier permeability and leakage space using dynamic Gd-DTPA scanning—a comparison of methods. Magnetic Resonance in Medicine **24** (1992) 174–176
3. Schmid, V.J., *et al.* : Statistical analysis of pharmacokinetical models in dynamic contrast-enhanced magnetic resonance imaging. In Duncan, J., Gerig, G., eds.: MICCAI 2005. (2005) 886–893
4. Twellmann, T., Lichte, O., Nattkemper, T.W.: An adaptive tissue characterization network for model-free visualization of dynamic contrast-enhanced magnetic resonance image data. IEEE Transactions on Medical Imaging **24** (2005) 1256–66
5. Gunn, R.N., Gunn, S.R., Turkheimer, F.E., Aston, J.A.D., Cunningham, V.J.: Positron emission tomography compartmental models: A basis pursuit strategy for kinetic modeling. J Cereb Blood Flow Metab **22** (2002) 1425–1439
6. Jerosch-Herold, M., Swingen, C., Seethamraju, R.: Myocardial blood flow quantification with MRI by model-independent deconvolution. Medical Physics **29** (2002) 886–897
7. Eilers, P., Marx, B.: Flexible smoothing with B-splines and penalties (with comments and rejoinder). Statistical Science **11** (1996) 89–121
8. Lang, S., Brezger, A.: Bayesian P-splines. Journal of Computational and Graphical Statistics **13** (2004) 183–212
9. Tofts, P., Kermode, A.: Measurement of the blood-brain barrier permeability and leakage space using dynamic MR imaging–1. Fundamental concepts. Magnetic Resonance in Medicine **17** (1991) 357–367
10. Gilks, W.R., Richardson, S., Spiegelhalter, D.J., eds.: Markov Chain Monte Carlo in Practice. Chapman & Hall, London (1996)
11. St.Lawrence, K.S., Lee, T.Y.: An adiabatic approximation to the tissue homogeneity model for water exchange in the brain: I. Theoratical derivation. Journal of Cerebral Blood Flow & Metabolism **18** (1998) 1365–1377
12. Buckley, D.L.: Uncertainty in the analysis of tracer kinetics using dynamic contrast-enhanced T_1-weighted MRI. Magnetic Resonance in Medicine **47** (2002) 601–6
13. Ah-See, M.L.W., *et al.*: Does vascular imaging with MRI predict response to neoadjuvant chemotherapy in primary breast cancer? Journal of Clinical Oncology (Meeting Abstracts) **22** (2004) 582

Segmentation of Brain MRI in Young Children

Maria Murgasova[1], Leigh Dyet[2], David Edwards[2], Mary Rutherford[2], Joseph V. Hajnal[2], and Daniel Rueckert[1]

[1] Visual Information Processing Group, Department of Computing, Imperial College London
[2] Department of Imaging Sciences, Faculty of Medicine, Imperial College London

Abstract. This paper describes an automatic tissue segmentation algorithm for brain MRI of young children. Existing segmentation methods developed for the adult brain do not take into account the specific tissue properties present in the brain MRI of young children. We examine the suitability of state-of-the-art methods developed for the adult brain when applied to the segmentation of the young child brain MRI. We develop a method of creation of a population-specific atlas from young children using a single manual segmentation. The method is based on non-linear propagation of the segmentation into population and subsequent affine alignment into a reference space and averaging. Using this approach we significantly improve the performance of the popular EM segmentation algorithm on brain MRI of young children.

1 Introduction

The problem of automatic segmentation of brain MRI has been extensively studied in past decade thanks to its applications in clinical studies of normal and diseased brain. Recent research of brain development in prematurely born children requires reliable and accurate automatic segmentation techniques for the neonatal and early childhood brain as well. Approximately 40 % of prematurely born children have cognitive, neurological or behavioural impairment [1] and the cerebral abnormalities underlying problems such as minor motor impairment or inattention still remain unclear. Accurate tissue quantification of the preterm brain in early childhood would allow us to study the growth of different structures and help to relate developmental outcomes to changes in volume and shape of anatomical structures.

In addition to noise, intensity non-uniformity, partial volume effect and natural tissue intensity variation present in the adult brain MRI, neonatal and infant brain MRI tissue intensities have even greater natural variation caused by ongoing process of myelination of white matter. This is most profound in neonates where white matter is extremely difficult to distinguish from grey matter on T1 weighted MRI sequences. By the age of 1 or 2 years, the majority of white matter has myelinated. However, the natural tissue intensity distribution overlap results in large misclassifications in central brain structures. This problem is further compounded by the fact that child brain differs significantly in size and

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 687–694, 2006.

shape from the adult brain, in particular in the central structures of the brain [2]. This causes problems for methods which rely on probabilistic atlases built from adults as spatial prior information.

In this paper we focus on the segmentation of brain MRI of 1 and 2-year-old children. We show how to bootstrap a probabilistic tissue atlas for a specific population based on registration of a single manual segmentation to a population and subsequent alignment to the space of a reference subject and averaging. We demonstrate that the results of the expectation-maximization method by Van Leemput *et al.* ([3], [4]) on young children can be significantly improved by using an atlas which appropriately captures the anatomical variability of the infant brain and compare the results with a simple version of a method inspired by Tasdizen *et al.* [5].

2 Related Work

To date the research in MRI brain segmentation has focused on adult and neonatal brains. The most popular class of tissue classification methods is based on expectation-maximization framework ([3], [4], [6], [7]) which iteratively interleaves calculation of intensity probability distributions and/or bias field with refinement of spatial probability maps. A mixture Gaussian distribution is assumed as a model for the voxel intensity probability distribution as it is easy to incorporate into EM framework. However, the real distribution can significantly differ from Gaussian for example due to partial volume effect, see Fig. 1. This can be overcome by estimating the real distributions [7] or by including a model for the partial volume effect in the EM framework [8].

Natural tissue intensity variation in different areas of the brain is another reason for misclassification. For example, grey matter in central brain structures is characterized by brighter intensities than cortical grey matter, and is therefore likely to be misclassified as white matter. Ren *et al* [9] reports such misclassification in adult brain using the EM-based segmentation method FAST [10]. However non-rigid registration-based label propagation [11] succeeds in this area of the brain as it does not assume any model for tissue intensities.

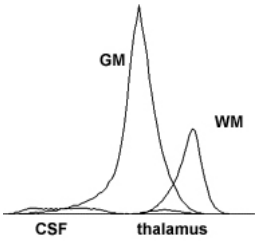


Fig. 1. Histogram of the tissue intensity probability distribution for white matter, grey matter, cerebrospinal fluid and thalamus (a deep grey matter structure) based on a manual segmentation of brain MRI from a 2-year-old child. Distributions of grey matter and CSF differ significantly from the Gaussian distribution due to partial volume effect. The intensity distribution of thalamus lies in the region of the overlap of the white and grey matter distributions.

Inevitable misclassifications resulting from overlaps in tissue intensity distributions can be partially avoided by spatially constraining the segmentation process with prior information in form of an aligned probabilistic atlas at each

iteration ([3], [4]); however this makes the method very sensitive to the correct alignment of the atlas with the image. This is very difficult for child brains which significantly differ from adult brains in size and shape [2]. To solve this problem, [12] and [13] suggest to warp the atlas to subject by non-rigid registration. In [14] Pohl reports difficulties with non-rigidly registering the template to image which is consistent with our own experience. Recent methods try to overcome this problem by iteratively warping a deformable atlas and refining the segmentation at the same time [15], [16].

An alternative non-parametric method for adult brain classification was developed by Cocosco *et al.* [17]. In this method the atlas is used only to select the training samples which are consequently used in kNN classification. This method is reported to deal well with anatomies which differ from the prior information. However, it cannot deal with the problem of tissue intensity distribution overlap in single-channel T1 data. Similarly, Tasdizen *et al.* [5] use the atlas only to initialize the estimation of multi-dimensional distribution of 7-neighbourhood.

The adult brain segmentation methods can be applied to the neonatal and infant brain when tuned to their specific properties. This was shown by Prastawa *et al.* [18] who used the EM approach ([3], [4]) with an atlas derived from single subject segmentation used as a prior, followed by non-parametric classification similar to [17] on neonatal data.

3 Methods

3.1 Expectation-Maximization Algorithm

For segmenting T1-weighted images from 1- and 2- year old children we have chosen the EM algorithm by Van Leemput *et al.*([3], [4]). We aligned the CCHMC pediatric brain atlas created from population of 5 to 9-year-old children [2] (Fig. 4c) with each subject's image. All images acquired on 1.0T HP MRI scanner were intensity corrected by the N3 software [19] in preprocessing step which proved to be sufficient as the bias field was not very strong. We have also acquired images on a 0.5T Apollo MRI scanner. In this case no bias field correction was necessary. This allowed us to study the EM method without the bias correction step and consequently better analyze its effect on the resulting segmentation.

The EM algorithm is a general technique for finding missing data based on observed data and maximum likelihood parameters estimates. In our case the observed data are the image intensities, the missing data are the labels and the parameters are the means and variances of the Gaussian distribution which is assumed for the intensity distribution of each tissue class. This is an iterative process which interleaves the calculation of posterior probabilities of each voxel belonging to each tissue class (white matter, grey matter, cerebrospinal fluid, other) - the expectation step, with maximum likelihood estimation of the Gaussian distribution parameters - the maximization step.

E-step:

$$p_{ik}^m = \frac{G(y_i, \mu_k^{m-1}, \sigma_k^{m-1}) p_{ik}^{atlas}}{\sum_j G(y_i, \mu_j^{m-1}, \sigma_j^{m-1}) p_{ij}^{atlas}}$$

M-step:

$$\mu_k^m = \frac{\sum_i p_{ik}^m y_i}{\sum_i p_{ik}^m} \quad \sigma_k^m = \frac{\sum_i p_{ik}^m (y_i - \mu_k^m)^2}{\sum_i p_{ik}^m}$$

where p_{ik}^m denotes the posterior probability of i^{th} voxel belonging to tissue k at m^{th} iteration, y_i is an intensity of i^{th} voxel and $G(.,.,.)$ denotes the Gaussian distribution with the mean μ_k^m and variance σ_k^m.

We found that this method significantly overestimates white matter in the central structures of the brain. This is most profound in the thalamus, a deep grey matter structure, where about 30% of voxels were misclassified as white matter (see Fig. 2 and Fig. 4b). The histogram based on the manual segmentation of a 2-year-old 1.0T data set (Fig. 1) shows that the voxel intensities of this structure lie in the region of the overlap of white and grey matter intensity distributions and therefore correct classification in this area can be achieved only by using correct prior information. However, prior information from the probabilistic atlas [2] overestimates white matter in this area as it does not describe the anatomy of the brain in early childhood adequately (Fig. 4c).

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

Fig. 2. Misclassification in thalamus: Substantial amount of deep grey matter classified as white. (a) MR image of a 2-year-old brain (b) EM segmentation using the adult atlas (c) EM segmentation using the 2-year-old child atlas (d) manual segmentation.

3.2 Bootstrapping the Prior Information for 1 and 2-Year-Old Brain

We obtained a manual segmentation of a single 2-year-old subject to create a probabilistic atlas of the 1 and 2-year-old brain. The manual segmentation was then transformed to 37 2-year-old subjects by performing non-rigid registration [11]. All 37 segmented subjects were then aligned with the reference subject by affine registration and the segmentations were averaged. The resulting probability maps describe the anatomy of the 2-year-old brain correctly when aligned with a subject by affine registration.

We also created a separate probabilistic atlas of the 1-year-old brain. First we non-rigidly warped the manual segmentation of a 2-year-old brain to the scan of

the same subject at 1 year to minimize the registration error. The same process of the atlas creation was then repeated with 34 1-year-old scans.

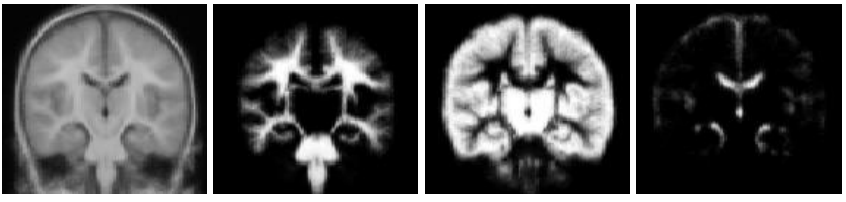
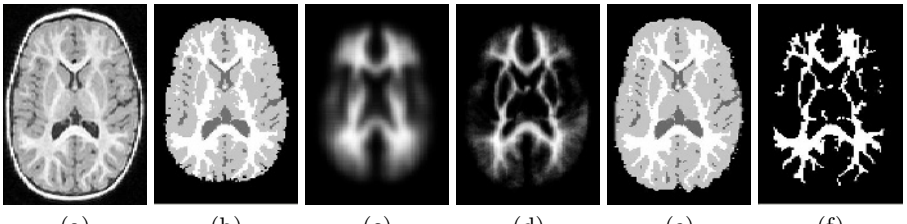


Fig. 3. The probabilistic atlas created from population of 2-year-old children

We were also able to use the single subject segmentation as the prior information by registering it non-rigidly to the image and subsequent blurring to accommodate anatomical variability.

Both of these methods resulted in significant improvement in segmentation of the central brain structures (Fig. 2 and 4). However, the white matter was still consistently overestimated which is caused by trying to separate the tissue classes based on single voxel intensity on global scale.



(a) (b) (c) (d) (e) (f)

Fig. 4. Improvement of segmentation using our new atlas: (a) MR image of a 2-year-old brain (b) segmentation using the 5 to 9-year-old CCHMC brain atlas (c) the 5 to 9-year-old CCHMC brain atlas (d) our 2-year-old brain atlas (e) segmentation using our 2-year-old brain atlas (f) manual segmentation

3.3 Local Distribution Estimation Method

In early childhood brain MRI exhibits significant natural intensity variability within the tissue classes. However, the boundaries are often visible to human observer. Therefore local estimation of tissue distributions can improve the segmentation. Inspired by [5] we implemented a local distribution estimation method which we have compared with the results of the EM approach in previous sections. A non-rigidly warped manual segmentation was used as prior information and the tissue intensity distributions were then estimated by Parzen windowing for each voxel separately after masking with local Gaussian kernel. Each voxel was then assigned to the tissue class with the highest probability. Even though

this approach improved the segmentation in small detail it was also less stable and sensitive to noise than the methods which estimate the tissue intensity distributions globally.

4 Validation and Results

4.1 Validation Using Manual Segmentation

First, we have validated our results on a single subject for which we obtained the manual segmentation. Validation was performed only in the brain area, excluding skull, brain stem and cerebellum. We segmented the image with our implementation of the EM method (section 3.1) using 3 different atlases as prior: our new probabilistic atlas based on population of 2-year-old children (2yr atlas), blurred manual segmentation (2yr seg) and the standard adult atlas obtained from SPM software [20] (adult atlas). For comparison we also segmented the image with EMS software [21] developed by Van Leemput. This software uses the SPM adult atlas, and unlike in our implementation, the bias correction is incorporated within the EM framework. We validated the approach with and without Markov Random Fields (MRF). Segmentation using local estimation of tissue parameters was evaluated as well. The manual segmentation was used as the ground truth. The correctness of the segmentation was calculated using Dice metric

$$D = \frac{2 * |T_{gt} \cap T_{seg}|}{|T_{gt}| + |T_{seg}|}$$

where T_{gt} and T_{seg} denote the set of samples of given tissue type in the ground truth and the automatic segmentation, respectively. The validation results (Fig. 1a) show that our new atlas improved the segmentation of both white and grey matter by 4% compared to EM segmentation with the adult atlas by both EMS and our implementation. The improvement is most significant in the thalamus where performance increased by 12% in comparison with EMS and by 26% compared to our implementation. Better performance of the EMS software in comparison with our implementation of EM with the adult atlas can be explained by the fact that the bias correction step partially compensates for natural intensity variation in the infant brain. When a blurred manual segmentation was used as prior information for EM, the performance slightly increased, but in this case it did not have to be non-linearly warped to the subject and therefore the results do not include the registration error. Overall performance of the local method was only marginally better than EMS due to sensitivity to noise and instability. However, it resulted in 94% correct segmentation of the thalamus. Again, the results do not include the registration error.

4.2 Two Step Validation

In the second experiment we compared our 3 methods on 37 2-year-old subjects. As we do not have the ground truth for any other subject we performed a two step segmentation. First, we segmented all the subjects by each of our methods and

Table 1. (a) Performance of the segmentation methods on a single 2-year-old subject (b) Two-step validation of the segmentation methods on 37 2-year-old subjects

	WM	GM	CSF	thalamus
EMS	0.844	0.883	0.694	0.779
EMS-MRF	0.855	0.891	0.616	0.799
EM adult atlas	0.847	0.879	0.672	0.664
EM 2yr atlas	0.885	0.922	0.784	0.923
EM 2yr seg	0.904	0.927	0.814	0.926
local	0.877	0.906	0.625	0.939

(a)

	WM	GM	CSF	thalamus
EM 2yr atlas	0.716	0.852	0.569	0.878
EM 2yr seg	0.872	0.907	0.740	0.847
local	0.855	0.809	0.404	0.916

(b)

the resulting probability maps were used as prior information for segmentation of the manually segmented subject. The segmentations were then compared with the manual segmentation using Dice metric (Fig. 1b). In this experiment the EM segmentation with non-rigidly warped blurred manual segmentation performed best overall whereas the local method was the most successful in the thalamus area.

5 Discussion and Conclusion

In this paper we have described the problem segmentation of the brain in early childhood with focus on segmentation of deep grey matter structures. Intensity based methods are likely to fail in this area as the voxel intensities are brighter than in cortical grey matter. This is caused by different composition of deep grey matter and presence of white matter tracts. However it is necessary to preserve the correct structure boundaries when performing white and grey matter classification. We showed that it is possible to substantially improve the results by using correct population-specific prior information. We also showed how to create such population specific probabilistic atlas. In future we would like to focus on development of a robust method which would perform tissue classification locally and incorporate iterative non-rigid warping of deformable atlas into segmentation process.

References

1. Marlow, N., Wolke, D., Bracewell, M.A., Samara, M., the EPICure Study Group: Neurologic and developmental disability at six years of age after extremely preterm birth. Pediatrics **115**(2) (2005) 286–94
2. Wilke, M., Schmithorst, V., Holland, S.: Normative pediatric brain data for spatial normalization and segmentation differs from standard adult data. Magnetic Resonance in Medicine **50**(4) (2003) 749–757
3. Leemput, K.V., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based tissue classification of MR images of the brain. IEEE Transactions on Medical Imaging **18**(10) (1999) 897–908

4. Leemput, K.V., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based bias field correction of MR images of the brain. IEEE Transactions on Medical Imaging **18**(10) (1999) 885–896
5. Tasdizen, T., Awate, S.P., Whitaker, R.T., Foster, N.L.: MRI tissue classification with neighborhood statistics: A nonparametric, entropy-minimizing approach. In: MICCAI '05, Part II. Lecture Notes in Computer Science (2005) 517–525
6. Ashburner, J.: Computational Neuroanatomy. PhD thesis, University College London (2000)
7. Wells III, W.M., Grimson, W.E.L., Kikinis, R., Jolesz, F.A.: Adaptive segmentation of MRI data. IEEE Transactions on Medical Imaging **15**(4) (1996) 429–442
8. Joshi, N., Brady, M.: A non-parametric mixture model for partial volume segmentation of MR images. In: British Machine Vision Conference. (2005)
9. Ren, J., Sneller, B., Rueckert, D., Hajnal, J., Hill, D.: A comparison of tissue type labelings of direct classification and segmentation propagation techniques for MR brain images. In: Proc. of Medical Image Understanding and Analysis '05. (2005) 55–58
10. Zhang, Y., Brady, M., Smith, S.: Segmentation of brain MR images through a hidden markov random field model and the expectation maximization algorithm. IEEE Transactions on Medical Imaging **20**(1) (2001) 45–57
11. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J.: Non-rigid registration using free-form deformations: Application to breast MR images. IEEE Transactions on Medical Imaging **18**(8) (1999) 712–721
12. Pohl, K., Wells, W., Guimond, A., Kasai, K., Shenton, M., Kikinis, R., Grimson, W., Warfield, S.: Incorporating non-rigid registration into expectation maximization algorithm to segment MR images. In: MICCAI '02. Volume 2488 of Lecture Notes in Computer Science. (2002) 564–572
13. D'Agostino, E., Maes, F., Vandermeulen, D., Suetens, P.: Non-rigid atlas-to-image registration by minimization of class-conditional image entropy. In: MICCAI '04, Part I. Lecture Notes in Computer Science (2004) 745–753
14. Pohl, K.: Prior Information for Brain Parcellation. PhD thesis, Massachusetts Institute of Technology (2005)
15. Ashburner, J., Friston, K.: Unified segmentation. NeuroImage **26** (2005) 839–851
16. Warfield, S.K., Kaus, M., Jolesz, F.A., Kikinis, R.: Adaptive template moderated spatially varying statistical classification. In: MICCAI '98. (1998) 431–438
17. Chris A. Cocosco, A.P.Z., Evans, A.C.: A fully automatic and robust brain MRI tissue classification method. In: Medical Image Analysis. Volume 7. (2003) 513–527
18. Prastawa, M., Gilmore, J.H., Lin, W., Gerig, G.: Automatic Segmentation of Neonatal Brain MRI. In: MICCAI '04, Part I. Lecture Notes in Computer Science (2004) 10–17
19. Sled, J.G.; Zijdenbos, A.E.A.: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Transactions on Medical Imaging **17**(1) (1998) 87–97
20. Statistical Parametric Mapping. (www.fil.ion.ucl.ac.uk/spm)
21. Expectation Maximization Segmentation. (www.medicalimagecomputing.com/EMS)

A Learning Based Algorithm for Automatic Extraction of the Cortical Sulci

Songfeng Zheng[1], Zhuowen Tu[2], Alan L. Yuille[1], Allan L. Reiss[3], Rebecca A. Dutton[2], Agatha D. Lee[2], Albert M. Galaburda[4], Paul M. Thompson[2], Ivo Dinov[1,2], and Arthur W. Toga[2]

[1] Department of Statistics, UCLA, Los Angeles, CA, USA
{sfzheng, ztu, yuille}@stat.ucla.edu
[2] Laboratory of Neuro Imaging, UCLA Medical School, Los Angeles, CA, USA
[3] School of Medicine, Stanford University, Stanford, CA, USA
[4] School of Medical, Harvard University, Cambridge, MA, USA

Abstract. This paper presents a learning based method for automatic extraction of the major cortical sulci from MRI volumes or extracted surfaces. Instead of using a few pre-defined rules such as the mean curvature properties, to detect the major sulci, the algorithm learns a discriminative model by selecting and combining features from a large pool of candidates. We used the Probabilistic Boosting Tree algorithm [16] to learn the model, which implicitly discovers and combines rules based on manually annotated sulci traced by neuroanatomists. The algorithm almost has no parameters to tune and is fast because of the adoption of integral volume and 3D Haar filters. For a given approximately registered MRI volume, the algorithm computes the probability of how likely it is that each voxel lies on a major sulcus curve. Dynamic programming is then applied to extract the curve based on the probability map and a shape prior. Because the algorithm can be applied to MRI volumes directly, there is no need to perform preprocessing such as tissue segmentation or mapping to a canonical space. The learning aspect makes the approach flexible and it also works on extracted cortical surfaces.

1 Introduction

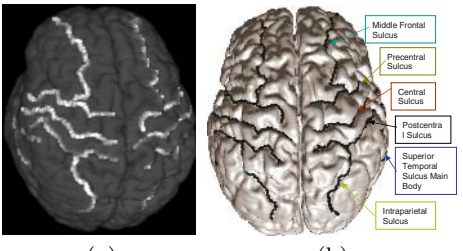
Cortical sulci are important structures of the brain. Reliably extracting major cortical sulci from MR images helps us to better understand the functionalities of the brain [7], facilitates many neuro studies for comparing different subjects[12], and assists other brain mapping tasks such as registration [5]. Fig. 1 shows several major cortical sulci: the Central Sulcus, the Postcentral Sulcus, the Superior Temporal Sulcus, the Intraparietal Sulcus, the Middle Frontal Sulcus and the Intraparietal Sulcus.

Cortical sulci lie on the valleys of cortical folds, and can be characterized by mathematical measures such as the mean curvatures [6, 1]. However, major cortical sulci have complicated geometric and photometric patterns and one often needs to use many protocols and high-level knowledge [9] to guide the manually labelling process. In general, automatic extraction of cortical sulci is

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 695–703, 2006.

a challenging task [12] due to their large intra-class variation and inter-class similarity across different subjects.

Most existing approaches for automatic cortical sulci detection work on the extracted cortical surfaces, which require a pre-processing stage to segment the tissue. Tao et al. [14] used global and local shape priors of sulcus curves on a canonical unit sphere to guide the extraction of the curves, but the method involved mapping the cortical surface to the unit sphere. Khaneja [6] used a dynamic programming approach to find the curves by minimizing an energy function, but this algorithm was not fully automatic since it required the starting and ending points of the sulci to be specified by hand. In [11], the patterns of different local folds are learned and they form a random graph using neural networks. The work by Rettmann [10] extracted sulcal regions using a watershed transformation method applied to cortical surfaces.



(a) (b)

Fig. 1. Examples of cortical sulci: (a) shows an MRI volume overlayed with several major cortical sulci, such as the Central Sulcus and the Postcentral Sulcus; (b) illustrates a corresponding extracted cortical surface with the same set of manually labelled sulcus curves

Our method is different from the above methods and it can be directly applied to either MRI volumes or extracted cortical surfaces. Instead of using a number of pre-defined rules/features, we learn/compute the likelihood of each voxel being on a sulcus curve based on a sub-volume ($15 \times 15 \times 15$) centered at this voxel. The *probabilistic boosting tree* (PBT) algorithm [16] is employed to select and combine hundreds of features from a large set of candidates to make an overall decision using a hierarchical structure. The candidate pool consists of around 5,000 features at three scales including intensity, gradients, curvatures, shape indices, locations, and 3D Haar filters etc.. A data-dependent prior is used to put geometric constraints on the curve. Similar to [6], we use a dynamic programming approach to combine the likelihood and shape prior. The algorithm is fully automatic, very general, and has almost no parameter to tune for different major sulcus curves. Moreover, it can be applied in other curve/object detection tasks in medical imaging.

We have a dataset of 40 volumes with several major cortical sulci manually labelled. We split them randomly into 15 training volumes and 25 testing volumes. Training is done both on the MRI volume and extracted cortical surfaces, and we compare the results on the testing images.

2 Problem Formulation

In this section, we give the problem formulation for 3D curve detection. For the rest of this paper, we use $\mathbf{V}$ to represent a 3D MRI volume. When a surface has been extracted, the algorithm works almost the same except for some minor modifications in computing gradients and curvature features.

For an input volume, $\mathbf{V}$, the task of cortical sulcus curve detection is to extract the 3D curve C:

$$C = \{\mathbf{r}_i, i = 0, \cdots, L\}$$

where L is the length of the curve, and $\mathbf{r}_i$ is the coordinates of the i^{th} point on the curve.

We define variable B as the background, i.e. the part of the volume not on the curve C. Clearly, $C \cap B = \emptyset$ and $\Lambda = C \cup B$ is the domain of the whole volume. We define a label y for each voxel $\mathbf{r}$, so that $y = +1$ if the voxel is on the curve C and $y = -1$ if it is on the background B. We then define a discriminative model $\hat{p}(C|\mathbf{V})$ to be of form:

$$-\log \hat{p}(C|\mathbf{V}) = E_1(C) + E_2(C) + E_3$$

where E_3 is a constant which does not depend on C (corresponding to the normalization constant for the distribution $\hat{p}(C|\mathbf{V})$). From the Bayesian point of view, the optimal curve C^* is the one maximizes $\hat{p}(C|\mathbf{V})$.

The term E_1 is given by:

$$\begin{aligned} E_1(C) &= -\sum_{\mathbf{r}\in B} \log p(\mathbf{V}(\mathbf{r}), y = -1|\mathbf{V}(N(\mathbf{r})/\mathbf{r})) - \sum_{\mathbf{r}\in C} \log p(\mathbf{V}(\mathbf{r}), y = +1|\mathbf{V}(N(\mathbf{r})/\mathbf{r})) \\ &= -\sum_{\mathbf{r}\in \Lambda} \log p(\mathbf{V}(\mathbf{r}), y = -1|\mathbf{V}(N(\mathbf{r})/\mathbf{r})) - \sum_{\mathbf{r}\in C} \log \frac{p(y = +1|\mathbf{V}(N(\mathbf{r})))}{p(y = -1|\mathbf{V}(N(\mathbf{r})))}. \end{aligned}$$

where $N(\mathbf{r})$ is the sub-volume centered at voxel $\mathbf{r}$; $\mathbf{V}(\cdot)$ is the intensity value(s) at the given voxel(s); $N(\mathbf{r})/\mathbf{r}$ includes all the voxels in the sub-volume except $\mathbf{r}$; $p(\mathbf{V}(\mathbf{r}), y|\mathbf{V}(N(\mathbf{r})/\mathbf{r}))$ is a conditional joint probability like pseudo-likelihood [2] in spirit. The first term in the above equation does not depend on C, therefore can be ignored. The probabilities $p(y = +1|\mathbf{V}(N(\mathbf{r})))$ and $p(y = -1|\mathbf{V}(N(\mathbf{r})))$ will be learned by PBT from manually labelled data.

The term E_2 is defined as:

$$E_2(C) = -\alpha L + \beta \sum_{i=0}^{L-1} \|\nabla \mathbf{V}(\mathbf{r}_{i+1}) - \nabla \mathbf{V}(\mathbf{r}_i)\|$$

where α and β are positive parameters to balance the importance of the corresponding terms. The motivation of introducing the first term is the observation that the sulcus curves are not very smooth, while the first term favors long curves, therefore, it discourages a curve being too straight. The second term of E_2 accumulates data dependent cues, i.e., we prefer that the intensity along the detected curve does not change too much.

In summary, maximizing the probability $\hat{p}(C|\mathbf{V})$ is equivalent to minimizing the energy function:

$$E(C) = -\sum_{\mathbf{r}\in C} \log \frac{p(y=+1|\mathbf{V}(N(\mathbf{r})))}{p(y=-1|\mathbf{V}(N(\mathbf{r})))} + E_2(C) \quad (1)$$

Here, the models capturing the appearances of foreground and background are combined in the discriminative probability ratio. Note that $p(y=+1|\mathbf{V}(N(\mathbf{r})))$ is the posterior probability of a voxel $\mathbf{r}$ belonging to the foreground (sulcus curve) given the sub-volume centered at $\mathbf{r}$. The second column in Fig. 3 shows such probability maps. The optimal curve C^* is the one that minimizes the above energy function $E(C)$.

We use dynamic programming (DP) to minimize $E(C)$ given by equation (1). DP is guaranteed to find the global minimum of $E(C)$, but requires starting and ending points (we refer to both as end points). We propose an adaptive method for selecting the end points, instead of choosing the end points by hand [6]. Our method uses the training data to measure the mean and covariance of the positions of the end points and then constrains the end points to lie within boxes centered on the means and with sides equal to twice the variance. We further localize the end points by requiring that $p(y=+1|\mathbf{V}(N(\mathbf{r})) > T$ where T is a learned threshold. Of these remaining points, we select the one that has the smallest value of $p(y=+1|\mathbf{V}(N(\mathbf{r}))$.

3 The Discriminative Models

Now the task is to learn and compute the discriminative probability $p(y|\mathbf{V}(N(\mathbf{r})))$ for each voxel $\mathbf{r}$. We adopt a new learning framework, probabilistic boosting tree [16], to learn complex discriminative models based on boosting algorithms [4].

For learning, we design a pool of around 5,000 features at three scales, including intensity, gradients, curvatures, shape indices, locations, and 3D Haar filters etc.. We used integral volume to compute the response of 3D Haar

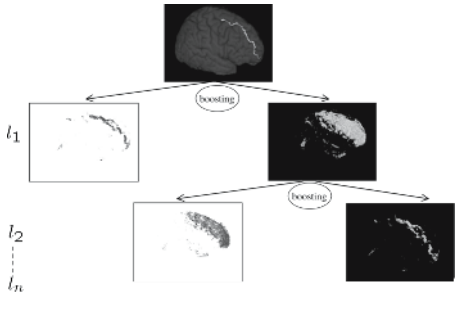


Fig. 2. Illustration of a boosting tree on a training volume: The left branch shows the probability of each voxel is not on the sulcus curve and the right branch shows the probability that each voxel is on the sulcus curve

filters. At each location (x_1, y_1, z_1), the integral volume is computed as $\int_0^{x_1}\int_0^{y_1}\int_0^{z_1} V(x,y,z)dxdydz$. Thus the computational cost for computing the response of 3D Haar filters is reduced significantly since every time we only need to sum up the values at the corners of the 3D Haar filter in the integral volume.

The tree is trained recursively: At the current node, the empirical distribution $\hat{q}(y)$ of the data is calculated, if the node is not pure (i.e., $\hat{q}(y)$ is not close to 0 or 1), a strong classifier is trained on the data. Each sample is then passed to the left and right subtree, weighted by $q(-1|\mathbf{V}(N(\mathbf{r})))$ and $q(+1|\mathbf{V}(N(\mathbf{r})))$, respectively, where $q(+1|\mathbf{V}(N(\mathbf{r})))$ is the probability that $\mathbf{V}(N(\mathbf{r}))$ is a positive sample, according to the strong classifier. Thus, the strong classifier at each node is not used to return the class of the sample but rather to assign the sample to the left or right subtree. Training proceeds recursively(see [16] for information about the basic algorithm). Fig. 2 shows the first two levels of a tree. PBT does training and testing in a divide-and-conquer manner and outputs the overall discriminative probability as:

$$\begin{aligned}\tilde{p}(y|\mathbf{V}(N(\mathbf{r}))) &= \sum_{l_1} \tilde{p}(y|l_1, \mathbf{V}(N(\mathbf{r})))q(l_1|\mathbf{V}(N(\mathbf{r})))\\ &= \sum_{l_1,l_2} \tilde{p}(y|l_2, l_1, \mathbf{V}(N(\mathbf{r})))q(l_2|l_1, \mathbf{V}(N(\mathbf{r})))q(l_1|\mathbf{V}(N(\mathbf{r})))\\ &= \sum_{l_1,\cdots,l_n} \tilde{p}(y|l_n, \cdots, l_1, \mathbf{V}(N(\mathbf{r}))), \cdots, q(l_2|l_1, \mathbf{V}(N(\mathbf{r})))q(l_1|\mathbf{V}(N(\mathbf{r})))\end{aligned}$$

where $l_i \in \{+1, -1\}$ is augmented hidden variable as shown in Fig. 2, indicating which branch is for this node: $l_i = -1$ and $l_i = +1$ point to the left and right branch, respectively. $q(l_i|l_{i-1}, \cdots, l_1, \mathbf{V}(N(\mathbf{r})))$ is the discriminative probability computed according to the boosting strong classifier learned at each node, and $\hat{q}(y|l_n, \cdots, l_1, \mathbf{V}(N(\mathbf{r})))$ is the empirical distribution at the leaf node.

4 Experiments

We have tested the proposed approach for extracting the Central and the Middle Frontal Sulci. In both cases, we used ground truth estimates of the Sulci location to train and test the algorithm. We also compared the performance for extracting the Central Sulcus when the cortical surface is available. The MR images, cortical surfaces and manually labelled landmark data are exactly the same as those in [13].

Learning the PBT took approximately eight hours (it is a function of the size of the training dataset); computing the posterior probability by PBT took about one minute per image; running the dynamic programming took around twenty seconds per image. The computer used has 2.4 GHz CPU and 1.0GB memory. Standard code optimization techniques can reduce these times significantly.

We have a dataset of 40 volumes of which we use 15 for training and the remained 25 for testing. In these volumes, the length of the sulci varies from 60-70 voxels. The PBT is learned on the 15 training volumes and the majority of features it selected to use were 3D Haar filters, though some curvature features

were also selected. Fig. 3 shows the results on some of the testing volumes for detecting the Central Sulcus and the Middle Frontal Sulcus on MRI volumes. The first column shows the input with manual labels superimposed and the second column shows the posterior probability of the sulci output by the PBT. Note that the probability maps have large responses around the correct position, but the maps are blurred, sometimes disconnected, and therefore are not sufficient to localize the sulcus directly. We then applied dynamic programming to minimize

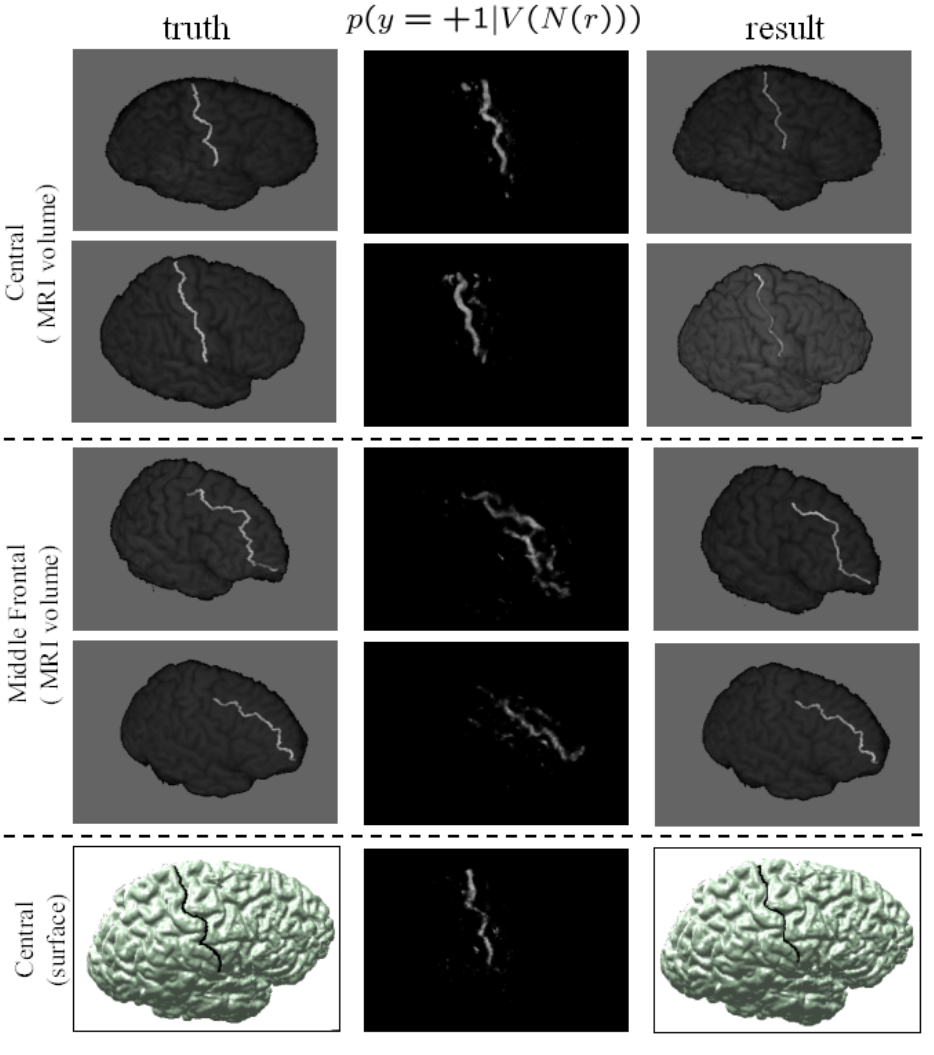


Fig. 3. Detection results on some of the testing images for the Central Sulcus on MRI, the Middle Frontal Sulcus on MRI, and the Central Sulcus on the cortical surfaces. The ground truth (left), the PBT posterior $p(y = +1|\mathbf{V}(N(\mathbf{r})))$ (middle), and the result of DP superimposed on the image (right).

the full energy function, equation (1). This resulted in a very clear estimate of the sulcus location, which are compared to the ground truth in Fig. 3.

Our method can also be applied to cases where only the cortical surface is available. This requires removing certain features from those considered in the case of volume data (i.e. we keep the 3D Haar filters, but remove intensity). We repeated the experiment on the Central Sulcus with the same training and testing dataset as before, and the result is also shown in Fig. 3.

To quantitatively evaluate the performance of our approach, we measured the distances between the estimated curves and the ground truth. We used the following distance measures:

$$H_{wor}(C,G) = \max_{c\in C} \min_{g\in G} |c-g|, \quad H_{av}(C,G) = \frac{1}{|C|}\sum_{c\in C} \min_{g\in G} |c-g|. \qquad (2)$$

Here C represents the curve found by DP and G is the ground truth. $H_{av}(C,G)$ gives the average distances from curve C to their closest points on curve G. By contrast, $H_{wor}(C,G)$ measures the worst case fit from curve C to curve G. For symmetry, we also consider $H_{av}(G,C)$.

Our evaluation results, as shown in table 1, show very good performance. In the table, $< \cdot >$ denotes the average over the dataset, for example, $< H_{av}(C,G) >= (1/N)\sum_{i=1}^{N} H_{av}(C_i,G_i)$, where N is the number of examples in the dataset. All the values of H_{av} are in the range of 3-5 voxels for both training and testing datasets. Observe that although the worst case measures are relatively big, the average distance is small, which suggests that some points occasionally have big offset while the overall curves still can be detected fairly accurately. The biggest errors occurred at the starting and ending points of the sulci. The testing errors are only slightly bigger than the training errors, which implies that our algorithm generalizes well.

Table 1. Error measures on 15 training and 25 testing images for the extracting of the Central Sulcus on volume, the Middle Frontal Sulcus on volume, and the Central Sulcus on surfaces. See text for notations. The unit of distance is the voxel.

Dataset	$< H_{av}(C,G) >$	$< H_{av}(G,C) >$	$< H_{wor}(C,G) >$
Testing (Central on volume)	2.74	3.46	7.53
Central (Central on volume)	3.76	4.22	8.56
Testing (Middle Frontal on volume)	4.26	4.59	12.04
Training (Middle Frontal on volume)	4.09	4.46	8.99
Testing (Central on surface)	2.79	3.07	9.48
Training (Central on surface)	2.44	2.88	8.44

5 Conclusions and Future Work

We presented a method for extracting cortical sulci, which uses a discriminative model including a discriminative term and a data dependent prior. The discriminative term was learned in a supervised way by the probabilistic boosting

tree algorithm, and dynamic programming was used to find the optimal solution by minimizing the discriminative energy function. A procedure was given for adaptively selecting the end points for dynamic programming.

The method was applied to extracting the Central Sulcus and the Middle Frontal Sulcus from raw MRI volume data. In both cases, the performance was very good when evaluated on both the training and testing data. The method was also applied to extracting the Central Sulcus when the cortical surface has already been extracted and the results were similar to those obtained directly from raw volume data. The biggest errors occurred at the end points of the sulci and might be due to ambiguity in the precise positions of these points (i.e. errors arising in the ground truth), or the procedure for estimating the starting and ending points. This is a topic for future research.

In this work, we designed a feature pool consists of about 5,000 candidates, most of which are 3D Haar filters. If effective features are used, we expect that the algorithm will achieve similar performance with fewer features. Our next step is to explore more sophisticated features for this particular medical image task.

Acknowledgement

This work was funded by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant U54 RR021813 entitled Center for Computational Biology (CCB).

References

1. A. Bartesaghi and G. Sapiro, "A System for the Generation of Curves on 3D Brain Images", *Human Brain Map.* 14:1-15 2001.
2. J. Besag, "Efficiency of pseudo-likelihood estimation for simple Gaussian fields", *Biometrika* 64: 616-618, 1977.
3. A. Caunce and C.J. Taylor, "Building 3D Sulcal Models Using Local geometry", *Medical Image Analysis*, 5 69-80, 2001.
4. Y. Freund and R. Schapire, "A Decision-Theoretic Generalization of On-line Learning and An Application to Boosting", *J. of Computer and Sys. Sci.*, 1997.
5. P. Hellier and C. Barillot, "Coupling Dense and Landmark-Based Approaches for Nonrigid Registration", *IEEE Trans. Med. Imaging*, vol. 22, no. 2, Feb. 2003
6. N. Khaneja, M.I. Miller, and U. Grenander, "Dynamic Programming Generation of Curves on Brain Surfaces", *PAMI*, vo.20, no. 11, 1998.
7. G. LeGoualher, E. Procyk, D.L. Collins, R. Venugopal, C. Barillot, A.C. Evans, "Automated extraction and variability analysis of sulcal neuroanatomy", *IEEE Trans. Med. Imaging* vo.18, no. 3, March 1999.
8. G. Lohmann, "Extracting Line Representations of Sulcal and Gyral Patterns in MR Images of the Hunam Brain", *IEEE Trans. on Medical Imag.*, vol.17, no.6, Dec. 1998.
9. M. Ono, S. Kubik, and S.D. Abernathey, "Atlas of the Cerebral Sulci", *New York:Thieme Medical*, 1990.

10. M.E. Rettmann, X. Han, C. Xu, J.L. Prince, "Automated sulcal segmentation using watersheds on the cortical surface", *Neuroimage*, 15(2):329-44, Feb. 2002.
11. D. Riviere, J.F. Mangin, D. Papadopoulos-Orfanos, J.M. Martinez, V. Frouin, and J. Regis, "Automatic Recognition of Cortical Sulci of the Human Brain Using A Congregation of Neural Networks", *Medical Image Analysis*, 77-92, 2002.
12. P.M. Thompson et al. "Mapping Cortical Change in Alzheimer's Disease, Brain Development, and Schizophrenia", *NeuroImage*, 23 Suppl 1:S2-18, September 2004.
13. P.M. Thompson, A.D. Lee, R.A. Dutton, J.A. Geaga,K.M. Hayashi KM, U. Bellugi, A.M. Galaburda, J.R. Korenberg, D.L. Mills, A.W. Toga, A.L. Reiss, "Abnormal Cortical Complexity and Thickness Profiles Mapped in Williams Syndrome", *J. of Neuroscience*, 25(18):, April 2005.
14. X. Tao, J.L. Prince, and C. Davatzikos, "Using a Statistical Shape Model to Extract Sulcal Curves on the Outer Cortex of the Human Brain", *IEEE Tran. on Medical Imag.*, vol.21, no. 5, May 2002.
15. J.P. Thirion and A. Gourdon, "Computing the Differential Characteristics of Isointensity Surfaces", *Computer Vision and Image Understanding*, vol. 61, no. 2, 1995.
16. Z. Tu, "Probabilistic Boosting-Tree: Learning Discriminative Models for Classification, Recognition, and Clustering", *Proc. of ICCV*, 2005.

Probabilistic Brain Atlas Encoding Using Bayesian Inference

Koen Van Leemput

Helsinki Medical Imaging Center, Helsinki University Central Hospital, Finland

Abstract. This paper addresses the problem of creating probabilistic brain atlases from manually labeled training data. We propose a general mesh-based atlas representation, and compare different atlas models by evaluating their posterior probabilities and the posterior probabilities of their parameters. Using such a Baysian framework, we show that the widely used "average" brain atlases constitute relatively poor priors, partly because they tend to overfit the training data, and partly because they do not allow to align corresponding anatomical features across datasets. We also demonstrate that much more powerful representations can be built using content-adaptive meshes that incorporate non-rigid deformation field models. We believe extracting optimal prior probability distributions from training data is crucial in light of the central role priors play in many automated brain MRI analysis techniques.

1 Introduction

The study of many neurodegenerative and psychiatric diseases benefits from fully-automated techniques that are able to reliably assign a neuroanatomical label to each voxel in MR images of the brain. In order to cope with the complex anatomy of the human brain, the large overlap in intensity characteristics between structures of interest, and the dependency of MRI intensities on the acquisition sequence used, state-of-the-art MRI brain labeling techniques rely on prior information in the form of *probabilistic atlases* [1, 2, 3, 4, 5, 6]. Typically, such atlases are created by voxel-wise averaging of neuroanatomical labels over a collection of manually labeled training datasets. In such an approach, the training datasets are first registered together, and the prior probability of each voxel being occupied by a particular structure is calculated as the relative frequency that structure occurred at that voxel across the training datasets.

While widely used, the quality of such "average" atlases as prior probability distributions has, to our knowledge, never been thoroughly investigated, and several open issues remain. In [6], for example, the authors are faced with the problem of creating a probabilistic atlas for newborn brain from only three training datasets. It is clear that, due to the enormous variability in cortical patterns across individuals, the average of three segmentations generalizes poorly to subjects not included in the training database, and the authors decided to blur their average atlas. But how much blurring should be used to obtain the "optimal" atlas? Is there still a need for blurring when more training data is used? Another

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 704–711, 2006.

question relates to the use of non-rigid registration techniques. Often atlases are constructed based on affine co-registrations of the training datasets, but are warped using non-rigid registration during the segmentation phase [7, 5, 2, 4]. While this inconsistency can be overcome by deforming the training data during the atlas construction phase itself [8, 9], a central question remains: how should the flexibility of the deformation models be chosen?

In this paper we explore ways to refine our atlas construction abilities beyond those currently available. From the discussion above, it is clear that finding good atlas models cannot simply be guided by how well a model describes the available training data: more complex models can always fit the training data better, leading to implausible, over-parameterized results. Rather, we will compare different models by evaluating their posterior probabilities and the posterior probabilities of their parameters. It is well-known that complex models are self-penalizing under Bayes' rule; we will show that a rigorous Bayesian approach is able to, among other things, provide quantitative answers to the questions raised above.

For the remainder of this paper, we will work with 2-dimensional image domains, with the understanding that the proposed techniques translate directly into 3 dimensions as well.

2 Mesh-Based Atlas Models

Let $L = \{l_i, i = 1, 2, \ldots, I\}$ be a manually labeled image with a total of I pixels, where $l_i \in \{1, 2, \ldots, K\}$ denotes the one of K possible labels assigned to pixel i. Partitioning the image domain D into T non-overlapping triangular elements, denoted by $D_t, t = 1, 2, \ldots T$, so that $D = \cup_{t=1}^{T} D_t$, we model the probability of having label l_i at pixel i by interpolation from the values at the nodes of the element D_t in which i falls:

$$p(l_i|\boldsymbol{\alpha}_{t,j}, \boldsymbol{x}_{t,j}) = \sum_{j=1}^{3} \alpha_{t,j}^{l_i} \varphi_{t,j}(\boldsymbol{x}_i). \tag{1}$$

In equation 1, $\boldsymbol{x}_i$ is the position of the the ith pixel, and $\boldsymbol{\alpha}_{t,j} = \{\alpha_{t,j}^1, \alpha_{t,j}^2, \ldots, \alpha_{t,j}^K\}$ denotes the set of label probabilities at the jth node of D_t. Furthermore, $\boldsymbol{x}_{t,j}$ is the position of the jth node of D_t, and $\varphi_{t,j}(\boldsymbol{x})$ is the linear interpolation basis function associated with this node.

For notational convenience, we will index the mesh nodes by $n = 1, 2, \ldots, N$ for the remainder of the paper, keeping in mind that each mesh node is typically shared among several triangles. Using this notation, equation 1 can be extended to cover the whole image domain D as follows:

$$p(l_i|\boldsymbol{\alpha}, \boldsymbol{x}, \mathcal{K}) = \sum_{n=1}^{N} \alpha_n^{l_i} \phi_n(\boldsymbol{x}_i) \tag{2}$$

where $\boldsymbol{\alpha} = \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_N\}$ and $\boldsymbol{x} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N\}$ is the set of the label probabilities and the positions of the mesh nodes, respectively, $\mathcal{K}$ denotes a

simplicial complex specifying the mesh connectivity, and $\phi_n(\boldsymbol{x})$ is the sum of the interpolation basis functions $\varphi_{t,j}(\boldsymbol{x})$ of the elements attached to node n.

Finally, assuming conditional independence of the labels between pixels given the mesh parameters, we have $p(L|\boldsymbol{\alpha}, \boldsymbol{x}, \mathcal{K}) = \prod_{i=1}^{I} p(l_i|\boldsymbol{\alpha}, \boldsymbol{x}, \mathcal{K})$ for the probability of seeing label image L.

2.1 First Level of Inference

Given certain training data in the form of M label images $L_m, m = 1, 2, \ldots, M$, and letting $\{\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^M\}$ denote the positions of the mesh nodes in each of the expert labelings, we may wish to infer what values of $\boldsymbol{\alpha}$ and $\{\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^M\}$ best fit the training data. To this end, we define a topology-preserving Markov random field prior on the position of the mesh nodes:

$$p(\boldsymbol{x}|\beta, \boldsymbol{x}^r, \mathcal{K}) \propto \exp(-\frac{U(\boldsymbol{x}|\boldsymbol{x}^r, \mathcal{K})}{\beta}),$$

$$\text{with} \qquad U(\boldsymbol{x}|\boldsymbol{x}^r, \mathcal{K}) = \sum_{t=1}^{T} -A_t^{\mathcal{K}}(\boldsymbol{x}^r) \log(A_t^{\mathcal{K}}(\boldsymbol{x})) \tag{3}$$

where $A_t^{\mathcal{K}}(\boldsymbol{x})$ denotes the area of the triangle t in a mesh with position $\boldsymbol{x}$, $\boldsymbol{x}^r$ is the most likely position of the mesh nodes (in the remainder called *reference* position), and the parameter β controls how far the mesh nodes can deviate from this reference position. Furthermore, having no specific prior knowledge about the values of the label probabilities $\boldsymbol{\alpha}$, we use a flat prior: $p(\boldsymbol{\alpha}) \propto 1$.

In a Bayesian setting, assessing the Maximum A Posteriori (MAP) parameters $\{\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{x}}^1, \ldots, \widehat{\boldsymbol{x}}^M\}$ involves minimizing

$$\sum_{m=1}^{M} -\log p(L_m|\boldsymbol{\alpha}, \boldsymbol{x}^m, \mathcal{K}) - \log p(\boldsymbol{x}^m|\beta, \boldsymbol{x}^r, \mathcal{K}). \tag{4}$$

We alternatively optimize the label probabilities in the mesh nodes $\boldsymbol{\alpha}$, keeping the position parameters fixed, and update each of the positions $\boldsymbol{x}^m$ while keeping the label probabilities fixed. Optimizing the positions is a registration process, bringing each of the training samples in spatial correspondence with the atlas. Since the gradient of equation 4 with respect to $\boldsymbol{x}^m$ is given by analytical expressions, we perform this registration by gradient descent. Assessing the optimal label probabilities in the mesh nodes for a given registration of the training samples can be done iteratively using an Expectation-Maximization (EM) algorithm. At each iteration, we calculate weights that associate each pixel in each example with each of the nodes attached to the triangle the pixel falls in

$$W_{i,n}^m = \frac{\alpha_n^{l_i^m} \phi_n^m(\boldsymbol{x}_i)}{\sum_{n'=1}^{N} \alpha_{n'}^{l_i^m} \phi_{n'}^m(\boldsymbol{x}_i)},$$

and update the parameters in each node n accordingly:

$$\alpha_n^k \quad \leftarrow \quad \frac{\sum_{m=1}^{M} \sum_{i=1}^{I} W_{i,n}^m \, \delta_{l_i^m, k}}{\sum_{m=1}^{M} \sum_{i=1}^{I} W_{i,n}^m} \qquad \forall n, k.$$

2.2 Second Level of Inference

The results of the atlas parameter estimation scheme described in section 2.1 depend heavily on the choice of the hyper-parameter β regulating the flexibility of the deformation fields. Having no prior knowledge regarding the "correct" value of β, we may assign it a flat prior. Using the Bayesian framework, we can then assess its MAP value $\widehat{\beta}$ by maximizing

$$p(L_1,\ldots,L_M|\beta,\boldsymbol{x}^r,\mathcal{K}) = \int_{\boldsymbol{\alpha}} \left(\prod_{m=1}^{M} p(L_m|\boldsymbol{\alpha},\beta,\boldsymbol{x}^r,\mathcal{K}) \right) p(\boldsymbol{\alpha}) \mathrm{d}\boldsymbol{\alpha} \tag{5}$$

$$\text{with} \quad p(L_m|\boldsymbol{\alpha},\beta,\boldsymbol{x}^r,\mathcal{K}) = \int_{\boldsymbol{x}^m} p(L_m|\boldsymbol{\alpha},\boldsymbol{x}^m,\mathcal{K}) p(\boldsymbol{x}^m|\beta,\boldsymbol{x}^r,\mathcal{K}) \mathrm{d}\boldsymbol{x}^m.$$

Assuming that $p(L_m|\boldsymbol{\alpha},\boldsymbol{x}^m,\mathcal{K})p(\boldsymbol{x}^m|\beta,\boldsymbol{x}^r,\mathcal{K})$ has a peak at a position $\boldsymbol{x}_\alpha^m$, we may approximate $p(L_m|\boldsymbol{\alpha},\beta,\boldsymbol{x}^r,\mathcal{K})$ using Laplace's method, i.e. by locally approximating the integrand by an unnormalized Gaussian. Using a similar Laplace approximation for the prior $p(\boldsymbol{x}^m|\beta,\boldsymbol{x}^r,\mathcal{K})$ in combination with the pseudo-likelihood approximation, and ignoring interdependencies between neighboring mesh nodes, we obtain[1]

$$p(L_m|\boldsymbol{\alpha},\beta,\boldsymbol{x}^r,\mathcal{K}) \;\simeq\; p(L_m|\boldsymbol{\alpha},\boldsymbol{x}_\alpha^m,\mathcal{K}) \cdot \prod_{n=1}^{N} O_n^m \tag{6}$$

$$\text{with} \quad O_n^m = \exp\Big(- \frac{U(\boldsymbol{x}_\alpha^m|\boldsymbol{x}^r,\mathcal{K}) - U(\boldsymbol{x}_\alpha^{m|n}|\boldsymbol{x}^r,\mathcal{K})}{\beta} \Big) \sqrt{\frac{\det(\boldsymbol{J}_n^m)}{\det(\boldsymbol{I}_n^m)}}$$

$$\text{where} \quad \boldsymbol{I}_n^m = D^2_{\boldsymbol{x}_n} \Big[- \log p(L_m|\boldsymbol{\alpha},\boldsymbol{x},\mathcal{K}) - \log p(\boldsymbol{x}|\beta,\boldsymbol{x}^r,\mathcal{K}) \Big] \Big|_{\boldsymbol{x}=\boldsymbol{x}_\alpha^m}$$

$$\text{and} \quad \boldsymbol{J}_n^m = D^2_{\boldsymbol{x}_n} \Big[- \log p(\boldsymbol{x}|\beta,\boldsymbol{x}^r,\mathcal{K}) \Big] \Big|_{\boldsymbol{x}=\boldsymbol{x}_\alpha^{m|n}}.$$

Here, $\boldsymbol{x}_\alpha^{m|n}$ denotes the set of mesh positions that is identical to $\boldsymbol{x}_\alpha^m$ except for the position of node n, which is replaced by the position to maximizes the prior $p(\boldsymbol{x}|\beta,\boldsymbol{x}^r,\mathcal{K})$ when the positions of all other mesh nodes are fixed to their value in $\boldsymbol{x}_\alpha^m$. Note that calculating this optimal node position, as well as evaluating the factors O_n^m, only involves those triangles that are directly attached to the node under investigation; we use Newton's method to carry out the actual optimization.

Plugging equation 6 into equation 5, approximating $\boldsymbol{x}_\alpha^m$ and the factors O_n^m by their values at $\boldsymbol{\alpha} = \widehat{\boldsymbol{\alpha}}$, introducing the EM algorithm's weights, and using Stirling's approximation, $x! \simeq x^x e^{-x}$, we finally obtain

$$p(L_1,\ldots,L_M|\beta,\boldsymbol{x}^r,\mathcal{K}) \simeq \prod_{m=1}^{M}\prod_{n=1}^{N} \widehat{O}_n^m \cdot \prod_{n=1}^{N} \frac{(K-1)!\,\widehat{N}_n!}{(\widehat{N}_n+K-1)!} \cdot \prod_{m=1}^{M} p(L_m|\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{x}}^m,\mathcal{K}) \tag{7}$$

[1] Here, D^2_θ denotes a matrix of second derivatives, or Hessian.

where $\widehat{N}_n = \left[\sum_{m=1}^{M}\sum_{i=1}^{I}\widehat{W}_{i,n}^m\right]$ denotes the total number of pixels associated with node n at the MAP parameters $\{\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{x}}^1, \ldots, \widehat{\boldsymbol{x}}^M\}$. Equipped with equation 7, we assess the MAP estimate $\widehat{\beta}$ using a line search algorithm.

2.3 Third Level of Inference

We have assumed so far that the connectivity $\mathcal{K}$ and the reference position $\boldsymbol{x}^r$ of the atlas mesh are known beforehand. Using the Bayesian framework, however, we can assign objective preferences to alternative models. Having no a priori reason to prefer one model over the other, we can rank alternatives based on their likelihood $p(L_1, \ldots, L_M|\boldsymbol{x}^r, \mathcal{K}) = \int_\beta p(L_1, \ldots, L_M|\beta, \boldsymbol{x}^r, \mathcal{K})\ p(\beta)\ \mathrm{d}\beta$, which can be approximated, using Laplace's method, by

$$\left(\sqrt{2\pi}\ p(\widehat{\beta}) \Big/ \sqrt{\frac{\partial^2}{\partial\beta^2}\Big[-\log p(L_1,\ldots,L_M|\beta,\boldsymbol{x}^r,\mathcal{K})\Big]\Big|_{\beta=\widehat{\beta}}}\right) \cdot\ p(L_1, \ldots, L_M|\widehat{\beta}, \boldsymbol{x}^r, \mathcal{K}).$$

The first factor is typically overwhelmed by the second one, so we will ignore it and compare alternative models based on equation 7, evaluated at the MAP estimate $\widehat{\beta}$.

While it is straightforward to compare models using equation 7, finding the mesh with connectivity and reference position that explicitly maximizes equation 7 is another matter. In this paper, we start from a dense regular triangular mesh, and use a mesh simplification technique borrowed from the computer graphics literature [10]. The technique yields increasingly coarse meshes by iteratively unifying two adjacent mesh nodes into a single node using a so-called edge collapse operation; each iteration removes the edge that yields the highest increase (or lowest decrease) in equation 7. For each edge collapse operation, we optimize the reference position $\boldsymbol{x}_n^r$ of the unified node n with respect to equation 7 using Powell's direction set. Finally, from the resulting hierarchy of meshes, we retain the one that yields the highest likelihood as evaluated by equation 7.

2.4 Description Length Interpretation

Given the central role of equation 7 in this paper, it is instructive to write it down in terms of the length, measured in bits, of the shortest message that communicates the training data without loss to a receiver when a certain model $\{\mathcal{K}, \boldsymbol{x}^r\}$ is used. Following Shannon theory, this length is $-\log_2 p(L_1, \ldots, L_M|\boldsymbol{x}^r, \mathcal{K})$, which we approximate by (equation 7)

$$-\sum_{m=1}^{M}\sum_{n=1}^{N}\log_2\widehat{O}_n^m \quad - \quad \sum_{n=1}^{N}\log_2\frac{(K-1)!\,\widehat{N}_n!}{(\widehat{N}_n+K-1)!} \quad - \quad \sum_{m=1}^{M}\log_2 p(L_m|\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{x}}^m, \mathcal{K}).$$

According to the three terms, the message can be imagined as being subdivided into three blocks. Prior to starting the communication, the transmitter estimates

the MAP estimates $\{\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{x}}^1, \ldots, \widehat{\boldsymbol{x}}^M\}$ as laid out in section 2.1. It then sends a message block that encodes, for each label image, the position of each mesh node (first term). Subsequently, a message block is sent that encodes the prior probabilities in each mesh node (second term), after which the actual data can be encoded using the model with the MAP parameters (third term).

3 Experiments

We evaluated the performance of our mesh-based atlas models on 2-D training data derived from publicly available manual annotations[2]. In a first experiment, we derived probabilistic atlases from 3 training datasets with 4 labels for three different model sub-groups (figure 1). The first sub-group uses a regular triangular mesh model and prevents the mesh nodes from moving away from the reference position $\boldsymbol{x}^r$ by setting $\beta = 0$ throughout. Within this sub-group we measured the description length as the resolution of the mesh, i.e. the distance between the mesh nodes, is varied, and retained the best model (top right in figure 1). The second sub-group also uses a regular triangular mesh model, but explicitly searches for the best β as the resolution of the mesh is varied (bottom left of figure 1). Finally, we used the mesh simplification procedure outlined in section 2.3 (bottom right of figure 1). The first sub-group did not perform well compared to the other groups: while it saves bits by not needing to encode the mesh node positions, it does not model the data very well, resulting in a long data message block. Clearly, letting β vary, as in the second sub-group, pays off, but an even shorter description length is obtained using the mesh simplification procedure.

In a second experiment, we added 15 more training datasets and searched again for the best resolution within the first sub-group (top part of figure 2). Comparing this with the result based on only three training datasets reveals that the number of mesh nodes has increased. Note that using a higher mesh resolution is akin to reducing the amount of blurring in the resulting atlas; Bayesian inference thus automatically and quantitatively determines the "correct" amount of blurring that should be applied. Also note that, while the mesh does have more nodes when eighteen training datasets are used, there are still far fewer mesh nodes than there are pixels (around 25 times fewer). The bottom part of figure 2 shows the atlas obtained by pixel-wise averaging, i.e. by setting the mesh resolution so high that each node corresponds to exactly one pixel. From the message length representation, it is clear that this is not a good model: there are far too many model parameters, resulting in a severely overfitted model.

In a final experiment, we searched for the best model in the same three sub-groups used before, using 9 training datasets containing 11 labels (figure 3). Again, the best atlas is obtained by mesh simplification.

[2] The Internet Brain Segmentation Repository, http://www.cma.mgh.harvard.edu/ibsr

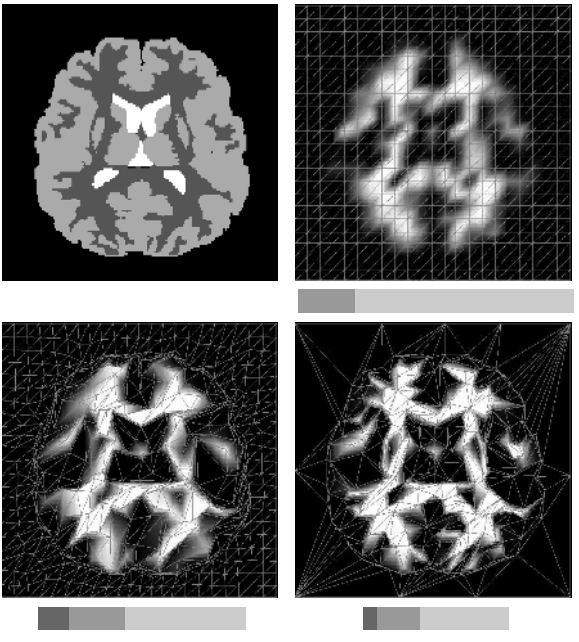


Fig. 1. Optimal probabilistic atlases constructed from three training datasets with four labels, for three subgroups of our mesh-based atlas models (see text for details). The top left figure depicts the first dataset of the training data; the two lower figures show atlases warped onto this dataset. Under each atlas is depicted a schematic view of the shortest message that encodes the training data: dark gray indicates the node position message block, intermediate gray represents the prior probabilities message block, and bright gray stands for the data message block.

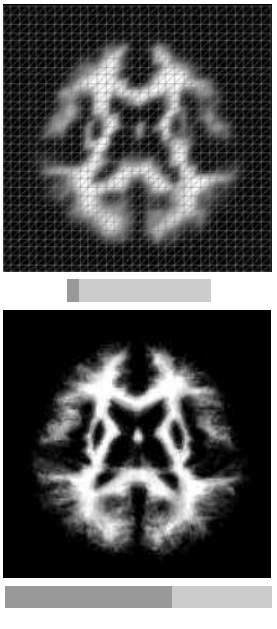


Fig. 2. "Average" atlas for eighteen training datasets at optimal resolution (top), and at pixel resolution (bottom)

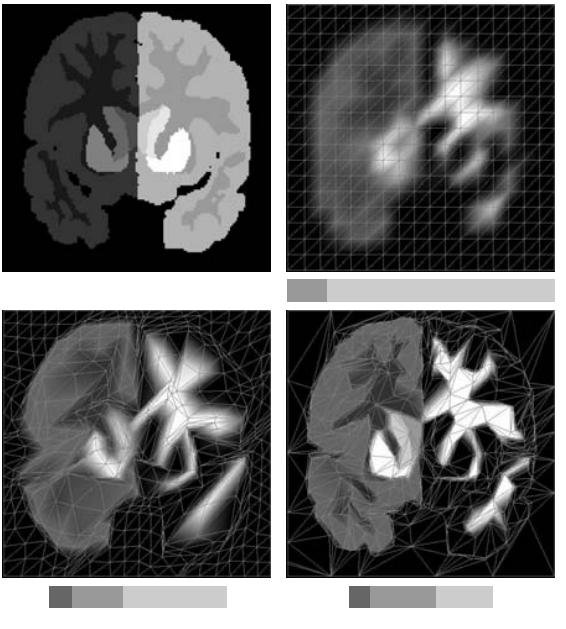


Fig. 3. Optimal probabilistic atlases based on nine images with eleven labels. The left side of the brain has been color-coded in the atlases for visualization purposes.

4 Discussion

While this paper concentrated on constructing prior probability distributions from training data, in practical segmentation scenarios the resulting atlases need to be aligned with the MRI data at hand before segmentation can commence. This requires that appropriate intensity distribution models are associated with each label. The gradient of the mesh node positions is then given in analytical form through equation 2, so that atlas-to-image registration is straight-forward to implement. Similar registration techniques that directly align priors with MR brain images have been described in [4, 7, 5].

Our future work will concentrate on implementing our atlas construction techniques in 3 dimensions, using tetrahedral rather than triangular meshes. We also plan to explore even more compact representations by explicitly encoding which subset of labels can occur at any given location throughout the brain.

References

1. K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens. Automated model-based tissue classification of MR images of the brain. *IEEE Transactions on Medical Imaging*, 18(10):897–908, October 1999.
2. K.M. Pohl, W.M. Wells, A. Guimond, K. Kasai, M.E. Shenton, R. Kikinis, W.E.L. Grimson, and S.K. Warfield. Incorporating non-rigid registration into Expectation Maximization algorithm to segment MR images. In *MICCAI 2002*, volume 2488 of *LNCS*, pages 564–571. Springer-Verlag, 2002.
3. A.P. Zijdenbos, R. Forghani, and A.C. Evans. Automatic "pipeline" analysis of 3-D MRI data for clinical trials: Application to multiple sclerosis. *IEEE Transactions on Medical Imaging*, 21(10):1280–1291, October 2002.
4. B. Fischl, D.H. Salat, A.J.W. van der Kouwe, N. Makris, F. Segonne, B.T. Quinn, and A.M. Dalea. Sequence-independent segmentation of magnetic resonance images. *NeuroImage*, 23:S69–S84, 2004.
5. J. Ashburner and K.J. Friston. Unified segmentation. *NeuroImage*, 26:839–851, 2005.
6. M. Prastawa, J.H. Gilmore, W. Lin, and G. Gerig. Automatic segmentation of MR images of the developing newborn brain. *Medical Image Analysis*, 9:457–466, 2005.
7. E. D'Agostino, F. Maes, D. Vandermeulen, and P. Suetens. Non-rigid atlas-to-image registration by minimization of class-conditional image entropy. In *MICCAI 2004*, volume 3216 of *LNCS*, pages 745–753. Springer-Verlag, 2004.
8. M. De Craene, A. du Bois d'Aische, B. Macq, and S.K. Warfield. Multi-subject registration for unbiased statistical atlas construction. In *MICCAI 2004*, volume 3216 of *LNCS*, pages 655–662. Springer-Verlag, 2004.
9. P. Lorenzen, M. Prastawa, B. Davis, G. Gerig, E. Bullitt, and S. Joshi. Multi-modal image set registration and atlas formation. *Medical Image Analysis*, 2006. In press.
10. H. Hoppe. Progressive meshes. In *ACM SIGGRAPH 1996*, pages 99–108, 1996.

Atlas Stratification*

Daniel J. Blezek and James V. Miller

GE Research, Niskayuna, NY 12309, USA
{blezek, millerjv}@research.ge.com

Abstract. The process of constructing an atlas typically involves selecting one individual from a sample on which to base or root the atlas. If the individual selected is far from the population mean, then the resulting atlas is biased towards this individual. This, in turn, can bias any inferences made with the atlas. Unbiased atlas construction addresses this issue by either basing the atlas on the individual which is the median of the sample or by an iterative technique whereby the atlas converges to the unknown population mean. In this paper, we explore the question of whether a single atlas is appropriate for a given sample or whether there is sufficient image based evidence from which we can infer multiple atlases, each constructed from a subset of the data. We refer to this process as *atlas stratification*. Essentially, we determine whether the sample, and hence the population, is multi-modal and is best represented by an atlas per mode. We use the mean shift algorithm to identify the modes of the sample and multidimensional scaling to visualize the clustering process.

1 Introduction

Atlas-based techniques have many applications in medical image analysis. Atlases take on many forms, ranging from an intensity image of the *average* subject to more detailed shape, intensity and functional models of specific structures. Atlases are used in basic research on population analysis, as guides in gross segmentation and seed point selection, as context in navigation tasks, and as models to overcome signal limitations and indistinct boundaries. Atlases may be based on a single individual or on a sample of a population. Atlases can be deterministic, where each region of space is associated with a single structure, or atlases can be probabilistic, where each region of space is assigned a likelihood of belonging to a variety of structures.

When atlases are constructed from a sample of a population, the imagery for the subjects in the sample are transformed into a common coordinate frame prior to consolidating their information. This step of *rooting* the atlas is common to both deterministic and probabilistic atlas construction. Establishing this common coordinate frame is a critical step that impacts the quality of the resulting atlas. A common practice is to select one subject from the sample on

* This work is part of the National Alliance for Medical Image Computing (NAMIC), funded by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant U54 EB005149. Information on the National Centers for Biomedical Computing can be obtained from http://nihroadmap.nih.gov/bioinformatics.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 712–719, 2006.

which to base the atlas. But if the selected subject is far from the population mean, the resulting atlas will be biased towards this individual. This, in turn, can bias any inferences made with the atlas. This issue has led to recent interest in *unbiased* atlas construction. Unbiased atlases can be constructed by searching for the *subject* closest to the population mean [1, 2] and rooting the atlas on that subject, or by searching for the *common coordinate frame* in the center of the population [3, 4, 5, 6, 7] and rooting the atlas on that coordinate frame.

Current atlas construction techniques are based on an assumption that the population is best described by a single atlas, treating the population as unimodal after transformation to the common coordinate frame. While this transformation may be non-rigid, and may therefore normalize away a portion of the inter-subject variability, substantial inter-subject variability may remain. Studying this remaining variability is the subject of population analysis. However, this same variability may render an atlas ineffective when used as a prior to combat signal limitations and indistinct boundaries. For these applications, variations beyond unimodal variations are particularly troubling.

In this paper, we explore the question of whether a population is best described by a single atlas or whether there is sufficient evidence to infer multiple atlases, each constructed from a subset of the data. We refer to this process as *atlas stratification*. We identify the modes in the population using a mean shift algorithm [8]. Each mode represents a subspace of the population which requires a unique atlas. In the process of identifying the modes, we determine which subjects should be used in constructing the atlas for each mode. The stratification process iterates between performing pairwise registrations of subjects and constructing atlases from subsets of subjects.

2 Mean Shift

Fukunaga and Hostetler introduced the mean shift algorithm [8] to estimate the gradient of a probability density function given a set of samples from the distribution. Using a hill climbing algorithm, this gradient estimate can be used to identify the modes of the underlying distribution. The mean shift algorithm has been used for clustering [8, 9], segmentation [10], and tracking [11].

Following the notation and derivation in [8], let $X_1, X_2, \ldots, X_N$ be a set of N iid. n-dimensional random vectors. The kernel density estimate of the underlying distribution is

$$\widehat{p}_N(X) = (Nh^n)^{-1} \sum_{j=1..N} k\left(\frac{1}{h}(X - X_j)\right) \tag{1}$$

where $k(X)$ is a scalar function satisfying the requirements for a kernel [12] and h is a parameter often referred to as the bandwidth [12]. If $k(X)$ is a differentiable function, the gradient of the density estimate is

$$\widehat{\nabla}_x p_N(X) = (Nh^n)^{-1} \sum_{j=1..N} \nabla_x k\left(\frac{1}{h}(X - X_j)\right) \tag{2}$$

where ∇_x is the gradient operator with respect to $x_1, x_2, \ldots, x_n$. A simple kernel of the form

$$k(X) = \begin{cases} c(1 - X^T X) & \text{if } X^T X \leq 1 \\ 0 & \text{if } X^T X > 1, \end{cases} \tag{3}$$

where c is a normalizing constant chosen to make the kernel integrate to one, satisfies the conditions for the density estimate to be asymptotically unbiased, consistent, and uniformly consistent [8]. Substituting this kernel into (2) yields

$$\widehat{\nabla}_x p_N(X) = (Nh^{n+2})^{-1} 2c \sum_{X_i \in S_h(X)} (X_i - X) \tag{4}$$

where $S_h(X)$ is a neighborhood with a radius equal to the bandwidth, h.

When $S_h(X)$ is small, $p_N(x)$ over the restricted domain of $S_h(X)$ is approximately uniform. The terms prior to the summation in (4) can be shown to be proportional to the density of an n-dimensional uniform distribution over $S_h(X)$. Therefore, we can approximate the normalized gradient (see [8] for details)

$$\frac{\widehat{\nabla}_x p_N(X)}{p_N(x)} \approx \frac{n+2}{h^2} M_h(X) \tag{5}$$

where

$$M_h(X) = \frac{1}{k} \sum_{X_i \in S_h(X)} (X_i - X)$$

is referred to as the sample mean shift, or simply the mean shift, and k is the number of samples in $S_h(X)$.

We can use this approximation to the normalized gradient to cluster samples X_j, $j = 1, 2, \ldots, N$, using the update equations

$$X_j^0 = X_j, \tag{6}$$

$$X_j^{t+1} = X_j^t + a \frac{\nabla_x p_N(X_j^t)}{p_N(X_j^t)}. \tag{7}$$

Using (5) and setting $a = \frac{h^2}{n+2}$, yields a simplified update equation

$$X_j^{t+1} = X_j^t + M_h(X_j). \tag{8}$$

This derivation of the mean shift is a k-nearest neighbor formulation, where the distance to the k^{th} nearest neighbor defines the bandwidth h.

3 Atlas Stratification

To apply the mean shift algorithm to the problem of atlas stratification, we consider the images for a subject as one sample. As such, each sample sits in a very high dimensional space (rows × cols × slices). We drop the standard

L_2 norm, used by the original mean shift algorithm to establish the nearest neighbor set, in favor of an information theoretic measure, specifically mutual information [13, 14, 15]. Other metrics and measures for comparing subjects for atlas stratification are possible. For instance, overlap metrics or shape similarity metrics applied to presegmented structures could be used in the mean shift framework.

Labeling the images for a subject as I_j and treating the N subjects as samples, let

$$I_j^0 = I_j \; j = 1, 2, \ldots, N.$$

At each iteration t, we align each pair of subjects using a Mattes mutual information registration algorithm [16]. The transformations that align each pair of subjects is T_{ij}^t. The mutual information values for each pair of subjects is $MI(T_{ij}^t \circ I_i^t, I_j^t)$. The nearest neighbor set for the mean shift iteration is the set

$$S_{d_k}(I_j^t) = \{I_i^t : MI(T_{ij}^t \circ I_i^t, I_j^t) \leq d_k\}$$

where d_k is k^{th} largest mutual information value to the image I_j^t. The mean shift is defined here as

$$M_h(I_j^t) = \frac{1}{k} \sum_{T_{ij}^t \circ I_i^t \in S_h(I_j^t)} (T_{ij}^t \circ I_i^t - I_j^t). \tag{9}$$

The samples are updated

$$I_j^{t+1} = I_j^t + M_h(I_j^t) \tag{10}$$

$$= I_j^t + \frac{1}{k} \sum_{T_{ij}^t \circ I_i^t \in S_h(I_j^t)} \left(T_{ij}^t \circ I_i^t - I_j^t\right) \tag{11}$$

where after each mean shift iteration, $t = t + 1$, the pairwise registrations are repeated using I_i^t and I_j^t from the previous iteration, producing new transformations T_{ij}^t and mutual information metric values $MI(T_{ij}^t \circ I_i^t, I_j^t)$. These transformations and metric values define the nearest neighbor sets for the next mean shift iteration. As the iterations of registrations and mean shift progress, the samples converge to the modes of the population.

4 Data

Our data was a random selection of 222 MR scans from the High Field MRI Studies of Neurodegenerative Disease conducted at the Albany Medical College's Neuroimaging Center. The scans were acquired on a 3T scanner (GE Medical Systems, Milwaukee WI). Mean age of the subjects was 74 years and ranged from 55-90 years. The scans were SPGR T1 weighted acquisitions with 15 deg flip angle, 12.1/5.2 TR/TE, 22 cm FOV, 2 mm slice thickness. In each scan, 96 coronal slices were acquired.

5 Results

The atlas stratification algorithm was implemented using the Insight Segmentation and Registration Toolkit [17]. Twenty iterations of atlas stratification were performed using mean shift bandwidths (k) of 7, 15, and 30 neighbors. At each iteration, 222x221 or 49062 volume registrations were performed followed by 222 averages of k volumes. The registrations were limited to estimating affine transformations. In total, the processing comprised over 3 million registrations and was distributed over a cluster of 500 computers.

5.1 Multidimensional Scaling

Multidimensional scaling is a cluster analysis technique that constructs a low-dimensional representation of a set of high dimensional samples given just the pairwise intersample distances [18]. Multidimensional scaling has previously been used in atlas construction by Park *et al* [2] to identify a subject close to the geometric mean of the population and rooting their atlas on that subject. Here, we use multi-dimensional scaling as a qualitative tool to visualize the mean shift clustering process. Figure 1 shows the results of multidimensional scaling applied to the intersubject distances after mean shift iterations 1, 2, 4, and 20 and for mean shift bandwidths of 7, 15, and 30 neighbors. Subjects in the same cluster at the last iteration are labeled with the same symbol. These symbol assignments are propagated back to earlier iterations to illustrate the clustering process. Note that multidimensional scaling produces a representation unique up to rotation/flip/permutation. Therefore, the visualizations across the iterations or down the bandwidth may vary in configuration by a rotation/flip/permutation.

To apply multidimensional scaling to the results of the mean shift algorithm, the pairwise mutual information metric values are converted to distances

$$D(I_i, I_j) = \alpha \left(\max_{I_k \in S} MI(I_k, I_j) \right) - MI(I_i, I_j) \tag{12}$$

where α is a small scale factor near unity to keep the distances distinct from zero. Since the mutual information values are the result of image registration, and these registrations are based on random sampling and gradient descent optimization, the mutual information values are not necessarily commutative. To combat this, the distance matrix $D(I_i, I_j)$ is made symmetric simply by

$$D(I_i, I_j) = \left(D(I_i, I_j) + D(I_j, I_i) \right) / 2.0 \tag{13}$$

and by prescribing the diagonal to be zero. These approximations are reasonable since the multidimensional scaling is only used to provide a qualitative assessment of the results of the mean shift algorithm.

Figure 1 shows the mean shift algorithm identifies multiple modes in the population. The number of modes being a function of the mean shift bandwidth. For a bandwidth of 30 neighbors, the mean shift algorithm produces 5 clusters containing 33, 39, 43, 47, and 60 subjects.

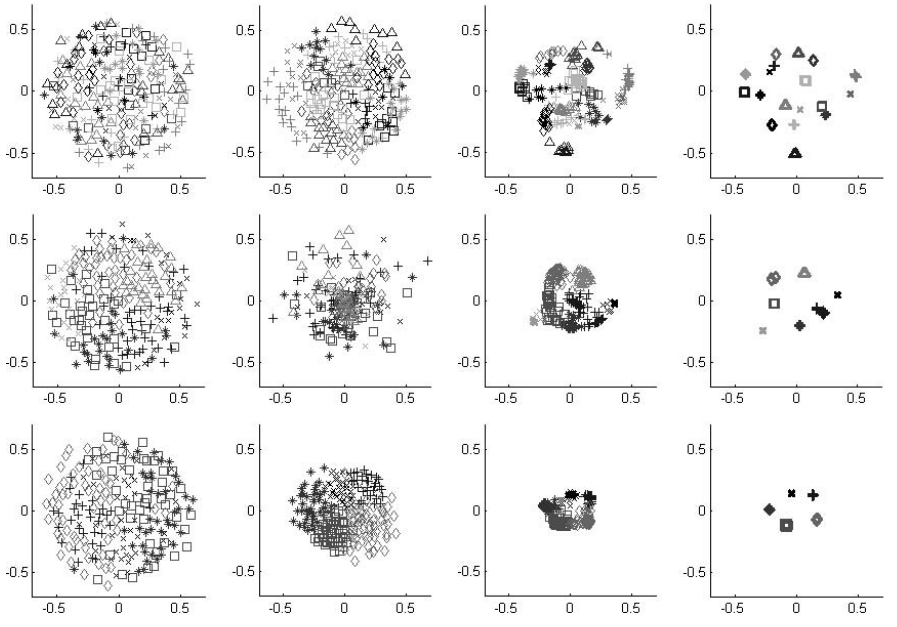


Fig. 1. Multidimensional scaling is an effective visualization tool for illustrating the progress of the mean shift clustering algorithm applied to the entire image database. The columns show multidimensional scaling applied to the output of mean shift iterations 1, 2, 4, and 20. The rows illustrate the effect of the mean shift bandwidth for nearest neighbor sizes 7, 15, and 30.

5.2 Atlas Modes

Figure 2 compares the atlases from two of the modes for a bandwidth of 30 neighbors. The atlas on the left is from a mode describing 43 subjects. The atlas on the right is from a mode describing 60 subjects. In each case, the atlas was

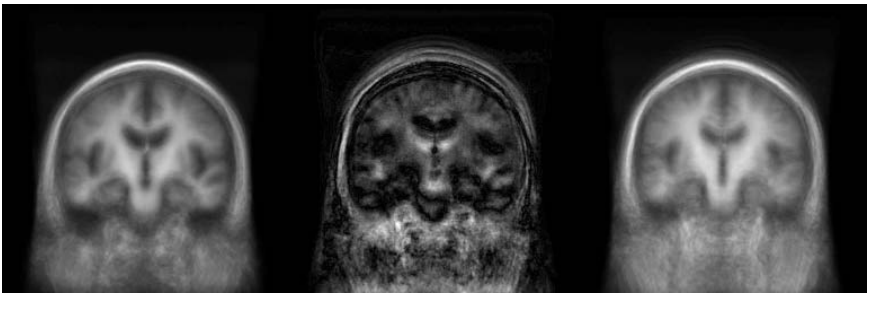


Fig. 2. Atlases from two modes (left, right) identified using a bandwidth of 30 neighbors and the difference between these two atlases (center)

constructed using only the 30 neighbors in the mean shift bandwidth. There are two interesting features to note. First, the atlas on the left has larger ventricles. However, the stratification process employed no explicit information on brain morphometry. Second, the atlas on the left is crisper. From this we infer subjects within this mode have less inter-subject variability than the subjects in the mode described by the atlas on the right (recall the same bandwidth, 30, was used to construct each atlas).

6 Conclusions

In this paper, we investigate *atlas stratification*, questioning whether a single atlas is appropriate for a given sample or whether there is evidence from which we can infer multiple atlases, each constructed from a subset of the data. We use the mean shift algorithm to search for modes in the population. If a population has multiple modes, the population should be described by multiple atlases to minimize bias. Our algorithm alternates between performing pairwise registrations of subjects and constructing atlases from subsets of the subjects. As the iterations of registrations and mean shift progress, the subjects converge to the modes of the population.

While the approach taken here is direct, it is not the only possible construction. For instance, the distance metric does not have to be based on mutual information. Overlap metrics or shape similarity metrics on presegmented structures could also be used in this mean shift framework. While our studies were based on an affine transform between subjects, higher order transformations and deformable registrations could be used. Mean shift formulations other than the nearest neighbor variant could also be used.

Aside from the above refinements, we've identified two areas of future research for atlas stratification. The first is a study of the algorithm itself, quantifying the differences between the atlases produced by atlas stratification. The second is a study of the algorithm in context, quantify an improvement in an atlas-based technique when multiple atlases are available. The latter will require a method to select the most appropriate atlas for a particular subject.

Acknowledgments

We would like to note our appreciation for the considerable efforts in collecting and managing the imagery used in this paper. Dr. John Schenck from GE Research and Dr. Earl Zimmerman from the Albany Medical College co-directed the collaboration on High Field MRI Studies of Neurodegenerative Disease as well as recruited volunteers for the study. Dave Henderson from GE Research provided data management and John Cowan from the Albany Medical College performed the subject scans.

References

1. S. Marsland, C. Twining, and C. Taylor. Groupwise non-rigid registration using polyharmonic clamped-plate splines. In *MICCAI*, pages 771–779, 2003.
2. H. Park, P.H. Bland, A.O. Hero, and C.R. Meyer. Least biased target selection in probabilistic atlas construction. In *MICCAI*, pages 419–426, 2005.
3. C. Studholme and V. Cardenas. A template free approach to volumetric spatial normalization of brain anatomy. *Pattern Recogn. Lett.*, 25(10):1191–1202, 2004.
4. K.K. Bhatia, J.V. Hajnal, B.K. Puri, A.D. Edwards, and D. Rueckert. Consistent groupwise non-rigid registration for atlas construction. In *ISBI*, pages 908–911, 2004.
5. M. De Craene, A. du Bois d'Aische, B. Macq, and S.K. Warfield. Multi-subject registration for unbiased statistical atlas construction. In *MICCAI*, pages 655–662, 2004.
6. P. Lorenzen, B. Davis, and S.C. Joshi. Unbiased atlas formation via large deformations metric mapping. In *MICCAI*, pages 411–418, 2005.
7. L. Zollei, E. Learned-Miller, E. Grimson, and W. Wells. Efficient population registration of 3d data. In *ICCV 2005*, 2005.
8. K. Fukunaga and L.D. Hostetler. The estimation of the gradient of a density function with applications in pattern recognition. *T-IT*, 21(1):32–40, 1975.
9. Y. Cheng. Mean shift, mode seeking and clustering. *T-PAMI*, 17(8):790–799, 1995.
10. D. Comaniciu and P. Meer. Mean shift: A robust approach toward featue space analysis. *T-PAMI*, 24(5):603–619, 2002.
11. D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean shift. In *CVPR 2000*, volume 2, pages 142–149, 2000.
12. B.W. Silverman. *Density estimation for statistics and data analysis.* Chapman and Hill, 1992.
13. W. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35–51, 1996.
14. A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multimodality medical image registration using information theory. In *Proc. 14th Int. Conf. Information Processing in Medical Imaging; Computational Imaging and Vision 3*, pages 263–274, 1995.
15. F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, Apr. 1997.
16. D. Mattes, D.R. Haynor, H. Vesselle, T.K. Lewellen, and W. Eubank. Pet-ct image registration in the chest using free-form deformations. *IEEE Transactions on Medical Imaging*, 22(1):120–128, Jan. 2003.
17. L. Ibanez, W. Schroeder, L. Ng, and J. Cates. *The ITK Software Guide.* Kitware, Inc. ISBN 1-930934-15-7, http://www.itk.org/ItkSoftwareGuide.pdf, second edition, 2005.
18. B.S. Everitt, S. Landau, and M. Leese. *Cluster Analysis.* Arnold, 2001.

Physiome Model Based State-Space Framework for Cardiac Kinematics Recovery*

Ken C.L. Wong[1], Heye Zhang[1], Huafeng Liu[3], and Pengcheng Shi[1,2]

[1] Department of Electronic and Computer Engineering,
Hong Kong University of Science and Technology, Hong Kong
{eewclken, eezhy, eeship}@ust.hk
[2] School of Biomedical Engineering, Southern Medical University, Guangzhou, China
[3] State Key Laboratory of Modern Optical Instrumentation,
Zhejiang University, Hanzhou, China
liuhf@zju.edu.cn

Abstract. In order to more reliably recover cardiac information from noise-corrupted patient-specific measurements, it is essential to employ meaningful *a priori* constraining models and adopt appropriate optimization criteria to couple the models with the measurements. While biomechanical models have been extensively used for myocardial motion recovery with encouraging results, the passive nature of such constraints limits their ability to fully count for the deformation caused by active forces of the myocytes. To overcome such limitations, we propose to adopt a *cardiac physiome model* as the prior constraint for heart motion analysis. The model is comprised of a cardiac electric wave propagation model, an electromechanical coupling model, and a biomechanical model, and thus more completely describes the macroscopic cardiac physiology. Embedded within a multiframe state-space framework, the uncertainties of the model and the patient-specific measurements are systematically dealt with to arrive at optimal estimates of the cardiac kinematics and possibly beyond. Experiments have been conducted on synthetic data and MR image sequences to illustrate its abilities and benefits.

1 Introduction

The goal of computational cardiac analysis is to objectively and accurately recover the parameters of various cardiac functions based on the specific measurements on the patients, such as those obtained from medical imaging, electrocardiograms, blood pressures, etc [1]. Nevertheless, these non-invasive, *in vivo* measurements can only provide either sparse, or gross, or projective observations in spatial and/or temporal domains, and are usually corrupted by noises of various sources. In consequence, analysis based solely on the measurements

* This work is supported in part by the China National Basic Research Program (973-2003CB716100), the National Natural Science Foundation of China (60403040), and the Hong Kong Research Grants Council (CERG-HKUST6151/03E).

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 720–727, 2006.

is ill-posed, and thus *a priori* models obtained from invasive or *in vitro* experiments, such as those from anatomy, biomechanics, or physiology, are necessary to constrain the inverse problems.

Study on human physiology has established that, at the beginning of each cardiac cycle, the pacemaker cells in the heart generate action potentials that are conducted to the whole heart through the conduction fibers which spread throughout the myocardium [1]. The myocardial contractile cells are then excited by the action potentials and contract according to the sequence of excitations, and the heart beats in a rhythmic motion. Thus, in order to properly model the cyclic dynamics of the heart, the cardiac electric wave propagation model (E model) that describes the spatiotemporal propagation pattern of the action potentials [2], the electromechanical coupling model (EM model) which transforms the excitation into myocardium contractile stresses [3], and the biomechanical model (BM model) which captures the deformation caused by the contractile stresses through the system dynamics [4] are all required. Since the integration of these three models gives a more complete macroscopic description of the physiological behavior of the heart, we call it the *cardiac physiome model* [5]. With *a priori* physiome models and various patient-specific measurements available, it becomes possible to establish a computational cardiac information analysis paradigm to recover the physiologically and clinically useful parameters such as the electrical propagation, ventricular geometry, cardiac kinematics, and myocardial material properties of the patient's heart.

In this paper, we address one aspect of the cardiac information analysis by proposing a kinematics recovery strategy that combines the physiome model with the multiframe filtering framework. Expressing the cardiac dynamics in a state-space representation, with the physiome model providing *a priori* predictive cardiac kinematics, patient specific kinematic estimates can be obtained in minimum-mean-square-error (MMSE) sense by updating the model predictions with measurements from images and other means. With the use of the physiome model, the predictions are more meaningful and potentially more accurate compared with the strategies which are only based on BM model and image-derived passive force [6]. Experiments have been performed on synthetic data to show the importance of utilizing the physiome model, and also performed on canine MR image sequences to show its possible practical applications.

2 Methodology

The physiome model, including the E, EM and BM components, provides prior predictions of cardiac kinematics, which are then updated by various measurement data to recover patient-specific cardiac parameters.

2.1 Electric Wave Propagation Model

After leaving the Purkinje fibers, the action potentials propagate throughout the myocardium through the myocytes. To model the electric wave propagation in the myocardium, because of the relative easiness of implementation, a simplified

reaction system based on FitzHugh-Nagumo model has been used as our model [7]. It is able to reproduce basic excitation propagation pattern through:

$$\begin{aligned} \frac{\partial u}{\partial t} &= c_1 f(u,v) + \bigtriangledown \cdot (D \bigtriangledown u) \\ \frac{\partial v}{\partial t} &= b(u - dv) \\ f(u,v) &= c_1 u(u-a)(1-u) - c_2 uv \end{aligned} \tag{1}$$

where u is the action potential that corresponds to the transmembrane voltage, v is the recovery variable, D is the diffusion tensor, $f(u,v)$ is the excitation term, and a, b, c_1, c_2 and d are parameters that define the shape of the action potential. The diffusion tensor and the parameters are constant in time but not necessary in space, with values adopted from [7].

2.2 Electromechanical Coupling Model

When a myocyte is excited by the action potential, it generates the active force for the contraction of the heart through *electromechanical coupling*. Although realistic complex models are available [4, 3], because of the computational efficiency, a simple ordinary differential equation proposed in [8] has been used as our electromechanical coupling model:

$$\frac{\partial \sigma_c}{\partial t} = \sigma_0 |u|_+ - |u| \sigma_c \tag{2}$$

where σ_c is a scalar related to the contraction stresses, σ_0 is a constant, and $|u|_+$ is the positive value of the normalized action potential u.

With σ_c obtained by solving Equation (2), the contraction stress tensor can be obtained by $\sigma_c f \otimes f$, with f the fiber orientation vector and $\otimes$ the tensor product. Then the active force R_c can be obtained as $R_c = \int_V \mathrm{div}(\sigma_c f \otimes f) dV$, with V the volume of the heart [9].

2.3 Biomechanical Model

The biomechanical model relates the active forces generated by the EM model with the resulted deformation of the myocardium. According to the anatomical literatures, the structure and the composition of the heart are very complicated. As a result, anisotropic material models are essential for the sensible description of myocardial characteristics.

Myocardium is assumed to be elastic. For both isotropic and anisotropic elastic materials, their stress-strain relation obeys the Hooke's Law:

$$\mathbf{S} = C\boldsymbol{\epsilon}, \tag{3}$$

where $\mathbf{S} = [S_{11}\ S_{22}\ S_{33}\ S_{12}\ S_{13}\ S_{23}]^T$ and $\boldsymbol{\epsilon} = [\epsilon_{11}\ \epsilon_{22}\ \epsilon_{33}\ \epsilon_{12}\ \epsilon_{13}\ \epsilon_{23}]^T$, with S_{ij} the components of the second Piola-Kirchhoff stress tensor, ϵ_{ij} the components of the Green-Lagrangian strain tensor, and C the stiffness matrix.

Because of the tubular shapes of the myocytes [4], we assume that the tissues of the myocardium are transversely isotropic. Let C_o be the 3D stiffness matrix of a point with 0^o polar and azimuthal angles as its fiber orientation:

$$C_o = \begin{bmatrix} \frac{1}{E_f} & -\frac{\nu_f}{E_{cf}} & -\frac{\nu_f}{E_{cf}} & 0 & 0 & 0 \\ -\frac{\nu_f}{E_{cf}} & \frac{1}{E_{cf}} & -\frac{\nu_{cf}}{E_{cf}} & 0 & 0 & 0 \\ -\frac{\nu_f}{E_{cf}} & -\frac{\nu_{cf}}{E_{cf}} & \frac{1}{E_{cf}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{G} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{G} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{2(1+\nu_{cf})}{E_{cf}} \end{bmatrix}^{-1}$$

where E_f, E_{cf}, ν_f, ν_{cf} are the Young's moduli and Poisson's ratios along and cross the fiber respectively, $G \approx E_f/(2(1+\nu_f))$ describes the shearing property.[1]

With C_o defined, the stiffness matrix at any point with known fiber orientation can be obtained using the tensor transformation:

$$C_{rot} = T^{-1} C_o R T R^{-1} \tag{4}$$

where T is the transformation matrix related to the fiber orientations, and R is the matrix responsible for the transformation between the strain tensor components and the engineering strain tensor components [6].

2.4 Cardiac System Dynamics Under Finite Deformation

The displacement-based total Lagrangian (TL) system dynamics of the heart under finite deformation can be written in the following form [6]:

$${}^t_0\mathbf{M}\ {}^{t+\Delta t}\ddot{\mathbf{U}} + {}^t_0\mathbf{C}\ {}^{t+\Delta t}\dot{\mathbf{U}} + {}^t_0\tilde{\mathbf{K}}\ \Delta\mathbf{U} = {}^{t+\Delta t}\mathbf{R}_c + {}^{t+\Delta t}\mathbf{R}_b - {}^t_0\mathbf{F} \tag{5}$$

where ${}^t_0\mathbf{M}$ is the mass matrix, ${}^t_0\mathbf{C}$ is the damping matrix, and ${}^t_0\tilde{\mathbf{K}}$ is the strain incremental stiffness matrix which contains the internal stresses as well as the material and deformation properties at time t. ${}^{t+\Delta t}\mathbf{R}_c$ is the force vector containing the contractile active force produced by the E and EM models, ${}^{t+\Delta t}\mathbf{R}_b$ is the force vector for enforcing boundary conditions, and ${}^t_0\mathbf{F}$ is the nodal force vector for finite deformation only and is related to the internal stresses at time t. ${}^{t+\Delta t}\ddot{\mathbf{U}}$, ${}^{t+\Delta t}\dot{\mathbf{U}}$ and $\Delta\mathbf{U}$ are the respective acceleration, velocity and incremental displacement vectors at time $t + \Delta t$.

Equation (5) is the governing cardiac system dynamics for finite deformation analysis, in which the material model is included through the embedding of material matrix C_{rot} into the matrix ${}^t_0\tilde{\mathbf{K}}$. In Equation (5), when the myocytes are excited and then provide the contractile forces ${}^{t+\Delta t}\mathbf{R}_c$ through Equations (1) and (2) to the system at time $t + \triangle t$, with the known structure and material property at time t, the incremental nodal displacements $\Delta\mathbf{U}$ (thus the cardiac kinematics at time $t + \triangle t$ caused by the active force) can be readily calculated, and they will be further updated by the imaging and other measurements to obtain the patient specific estimates with the use of multiframe filtering.

[1] If E_{cf} and E_f are equal, then C_o reduces to the stiffness matrix for an isotropic material.

2.5 Optimal Multiframe Kinematics Estimation: Coupling of Physiome Model and Patient's Measurements

The system dynamics of Equation (5) can be solved by applying the Newmark method and the Newton-Raphson iteration, thus providing the deterministic kinematic predictions solely based on the physiome model. Nevertheless, the parameters of the physiome model, such as the fiber orientations and the Young's moduli, may not coincide with the specific physiological properties of a particular patient's heart, and thus introduces the *system uncertainties*. On the other hand, patient-specific measurements are typically corrupted by noises (*observation errors*), so they cannot be fully trusted when fed into the system. In order to properly use the patient measurements and compromise with the prior physiome model, a statistical framework is required to couple them together.

Equation (5) can be rewritten as a state-space updating equation [6]:

$$x(t + \Delta t) = A(t)x(t) + w(t + \Delta t) \tag{6}$$

where $x(t + \Delta t) = {}^{t+\Delta t}\mathbf{U}$ and $x(t) = {}^{t}\mathbf{U}$ are the state vectors, containing the displacements at time t and time $t + \triangle t$. $A(t)$ is the *transition matrix* relating $x(t+\triangle t)$ with $x(t)$, and is comprised by the ${}^{t}_{0}\mathbf{M}$, ${}^{t}_{0}\mathbf{C}$ and ${}^{t}_{0}\tilde{\mathbf{K}}$ matrices of Equation (5). $w(t + \Delta t)$ is the *input vector* which contains the force inputs including the contractile force obtained from the physiome model. Also considering the zero-mean, additive, white *process noise* $v(t)$ ($E[v(t)] = 0$, $E[v(t)v(s)'] = Q_v(t)\delta_{ts}$), Equation (6) becomes a TL-updated state-space equation which performs nonlinear state prediction (displacement) using the Newton-Raphson iteration scheme:

$$x(t + \Delta t) = A(t)x(t) + w(t + \Delta t) + v(t + \Delta t) \tag{7}$$

and $x(t + \Delta t)$ becomes statistical representation with system uncertainties.

In order to apply the patient's measurements to the physiome model, the measurements are defined to be a subset of the state vector corrupted by noises. Assuming zero-mean, additive, and white noise $e(t)$, ($E[e(t)] = 0$, $E[e(t)e(s)'] = R_e(t)\delta_{ts}$), the measurement equation becomes the following:

$$y(t + \Delta t) = Hx(t + \Delta t) + e(t + \Delta t), \tag{8}$$

where H is a known *measurement matrix* and $y(t + \Delta t)$ contains the image-derived, patient-specific displacements.

With these statistical state-space equations (Equations (7) and (8)) defined, using the multiframe filtering procedure described in [6], the prior physiome kinematic predictions can be updated by measurements to provide the optimal patient-specific estimates.

3 Experimental Results

Synthetic Data. The canine heart model of the University of Auckland [2], which provides both the cardiac geometry and fiber architecture obtained through

[2] Bioengineering Institute, University of Auckland, http://www.bioeng.auckland.ac.nz

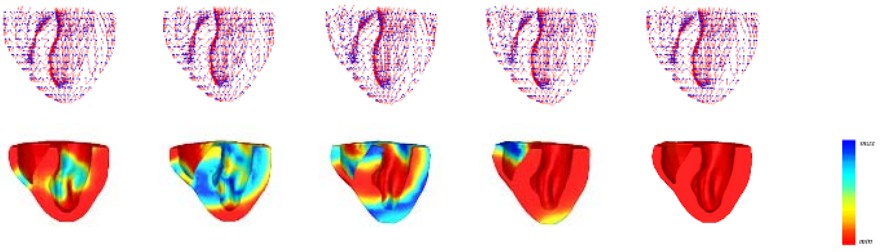


Fig. 1. Synthetic data generated by applying the physiome model (E, EM and BM models) on the canine heart model of the University of Auckland. Top: deformed geometries (blue) with fiber orientations (red) (frame #5, #15, #25, #35, #45). Bottom: action potential propagation (frame #5, #15, #25, #35, #45).

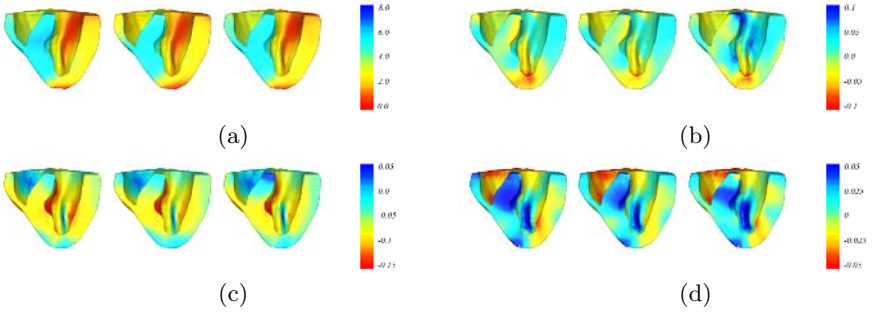


Fig. 2. Kinematics recovery results on synthetic data in cylindrical coordinate system (frame #25, with respect to the first frame, the longitudinal-axis is the long-axis of the left ventricle). (a) Displacement magnitude (U) maps. (b) Radial strain (ϵ_{rr}) maps. (c) Circumferential strain (ϵ_{cc}) maps. (d) Radial-circumferential strain (ϵ_{rc}) maps. Left to right: ground truth, multiframe physiome model estimation, multiframe passive biomechanical model estimation.

anatomical experiments, has been used to generate the synthetic data. The physiome model (E, EM, and BM models) was applied on the heart model to provide the ground truth (see Fig. 1), and 50 frames of one cardiac cycle were obtained. The displacements on the epicardium and endocardia of every frame were extracted, and noises with 5dB SNR were added to provide the noisy measurements (y in Equation (8)), which were treated as the patient-specific motion tracking results from the cardiac images.

The kinematic parameters of the synthetic heart were recovered by the proposed multiframe physiome model and the multiframe passive biomechanical model [6]. In Fig. 2, for frame #25 as an example, it can be seen that the strains estimated using our proposed multiframe physiome model are better, especially for those parts further from the measurements, such as the septum between the left and right ventricles. This is because the roles of the models are more im-

Table 1. Deviations of the estimates from the ground truth (synthetic data).

Experiment	Multiframe Physiome Estimation	Multiframe Passive Estimation
ϵ_{rr}	0.0085±0.0126	0.0228±0.0337
ϵ_{cc}	0.0091±0.0077	0.0120±0.0150
ϵ_{zz}	0.0061±0.0064	0.0101±0.0122
ϵ_{rc}	0.0059±0.0075	0.0104±0.0123
ϵ_{cz}	0.0048±0.0062	0.0066±0.0079
ϵ_{zr}	0.0059±0.0074	0.0098±0.0136
U	0.3792±0.2976	0.2266±0.2621

portant for those parts without measurements. In consequence, more realistic models give more accurate estimations.

The tabulated numerical results on the deviations from the ground truth of these experiments (all 50 frames) are shown in Table.1.

Canine MRI Data. Experiments have also been conducted on canine cardiac MR image sequences to show the possible applicability of the multiframe physiome framework to clinical data. The cardiac image sequence of a normal mongrel dog consists of 16 frames over the cardiac cycle, where each 3D image frame contains 16 image slices, with 5mm inter-slice spacing, in-plane resolution of 1.64mm/pixel, and temporal resolution of 40msec/frame.

In the experiments, the canine cardiac boundaries and initial motion correspondences of the left ventricle were obtained using a simultaneous segmentation and motion tracking algorithm based on the active region model to provide the

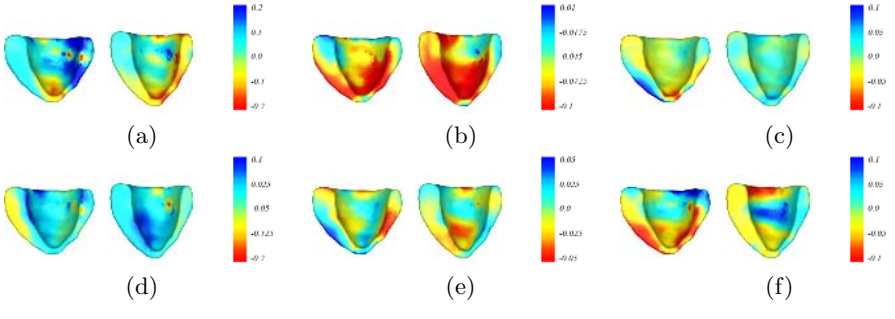


Fig. 3. Estimated kinematics of canine MRI data in cylindrical coordinate system (frame #8, with respect to the first frame, the longitudinal-axis is the long-axis of the left ventricle). (a) Radial strain (ϵ_{rr}) maps. (b) Circumferential strain (ϵ_{cc}) maps. (c) Longitudinal strain (ϵ_{zz}) maps. (d) Radial-circumferential strain (ϵ_{rc}) maps. (e) Circumferential-longitudinal strain (ϵ_{cz}) maps. (f) Longitudinal-radial strain (ϵ_{zr}) maps. Left to right: multiframe physiome model estimation, multiframe passive biomechanical model estimation.

measurements [10]. The fiber architecture of the heart model of the University of Auckland was mapped to the canine heart, using the iterative closest points algorithm [11] followed by local registration using deformable BM model.

The kinematic estimates obtained using the proposed multiframe physiome model and the multiframe passive biomechanical model are shown in Fig. 3. Since there is no ground truth provided, numerical results are not available for comparison. Nevertheless, it has been observed that the motion tracking results (the measurements) have nearly no deformation along the long-axis of the heart (the z-axis), and this is unrealistic to a normal heart. It can be seen that, especially in Fig. 3(c), the estimates obtained through the multiframe passive biomechanical model inherit the same z-axis motion from the measurements. In contrast, the estimates obtained using our proposed algorithm is more reasonable because the physiome model gives more realistic predictions to compromise with the unrealistic measurements.

References

1. Braunwald, E., Zipes, D., Libby, P.: Heart Disease: A Textbook of Cardiovascular Medicine. 6th edn. W.B. Saunders Company (2001)
2. Knudsen, Z., Holden, A., Brindley, J.: Qualitative modelling of mechano-electrical feedback in a ventricular cell. Bulletin of Mathematical Biology **6** (1997)
3. Nash, M.: Mechanics and Material Properties of the Heart using an Anatomically Accurate Mathematical Model. PhD thesis, University of Auckland (1998)
4. Glass, L., Hunter, P., McCulloch, A., eds.: Theory of Heart : Biomechanics, Biophysics, and Nonlinear Dynamics of Cardiac Function. Springer-Verlag (1991)
5. McCulloch, A., Bassingthwaighte, J., Hunter, P., Noble, D.: Computational biology of the heart: From structure to function. Progress in Biophysics and Molecular Biology **69** (1998) 153–155
6. Wong, K.C., Shi, P.: Finite deformation guided nonlinear filtering for multiframe cardiac motion analysis. In: International Conference on Medical Image Computing and Computer Assisted Intervention. (2004) 867–874
7. Rogers, J., McCulloch, A.: A collocation-Galerkin finite element model of cardiac action potential propagation. IEEE Transactions on Biomedical Engineering **41** (1994) 743–757
8. Sermesant, M., Coudière, Y., Delingette, H., Ayache, N.: Progress towards an electro-mechanical model of the heart for cardiac image analysis. In: IEEE International Symposium on Biomedical Imaging. (2002) 10–14
9. Ayache, N., Chapelle, D., Clément, F., Coudière, Y., Delingette, H., Désidéri, J., Sermesant, M., Sorine, M., Urquiza, J.: Towards model-based estimation of the cardiac electro-mechanical activity from ECG signals and ultrasound images. In: Functional Imaging and Modeling of the Heart. (2001) 120–127
10. Wong, L., Liu, H., Sinusas, A., Shi, P.: Spatio-temporal active region model for simultaneous segmentation and motion estimation of the whole heart. In: IEEE Workshop on Variational, Geometric and Level Set Methods in Computer Vision. (2003) 193–200
11. Besl, P., McKay, H.: A method for registration of 3-D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence **14** (1992) 239–256

Automated Detection of Left Ventricle in 4D MR Images: Experience from a Large Study

Xiang Lin[1], Brett R. Cowan[2], and Alistair A. Young[1]

[1] Bioengineering Institute, University of Auckland, New Zealand
[2] Center for Advanced MRI, University of Auckland, New Zealand
{x.lin, b.cowan, a.young}@auckland.ac.nz

Abstract. We present a fully automated method to estimate the location and orientation of the left ventricle (LV) in four-dimensional (4D) cardiac magnetic resonance (CMR) images without any user input. The method is based on low-level image processing techniques incorporating anatomical knowledge and is able to provide rapid, robust feedback for automated scan planning or further processing. The method relies on a novel combination of temporal Fourier analysis of image cines with simple contour detection to achieve a fast localization of the heart. Quantitative validation was performed using 4D CMR datasets from 330 patients (54024 images) with a range of cardiac and vascular disease by comparing manual location with the automatic results. The method failed on one case, and showed average bias and precision of under 5mm in apical, mid-ventricular and basal slices in the remaining 329. The errors in automatic orientation were similar to the errors in scan planning as performed by experienced technicians.

1 Introduction

We investigated methods for the robust, accurate and fully automatic identification of heart location and orientation from CMR examinations. The method is targeted for clinical applications and must therefore be fast, efficient and reliable. It should be able to return the location, orientation and approximate contours of the LV in the absence of any user input. The method is expected to have two important applications. Firstly, the detected LV contours could be used as input to higher level segmentation methods and deformable model based analysis. Secondly, the method can be used to speed up the acquisition time by facilitating the fully automatic planning of CMR examinations.

Segmentation of the LV in CMR images is important for quantitative assessment of the function of the heart and many automatic approaches at different image processing levels have been proposed for this problem. Low level techniques depending on the local intensity characteristics are fast but lack robustness. *A priori* knowledge can be incorporated in deformable model-based methods, however, the efficiency and robustness of the methods is heavily dependent on the initial contours or models. Most semi-automatic methods like [1] require manual initialization. Fully automatic algorithms can be found but many of these

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 728–735, 2006.

are computation-intensive [2] or lack of a wide range of clinical validation [3]. One [4] has been validated in 121 cases but assumes that the location of the heart is approximately at the center of the MR image. Specialized methods have also been proposed for tagged images [5] or perfusion images [6], however their application to patients with a wide range of clinical disease is unclear.

Automated CMR image planning has been proposed for speeding up scan time [7, 8]. The core requirement is for a fast and accurate calculation of the three-dimensional (3D) position and orientation of the LV. A deformable template based method was addressed in [7], which estimated the LV axis by fitting many feature points of major organs in the thorax to the localizer images. However, the method was computation-intensive. To avoid this problem, [8] proposed another method which employed *a priori* knowledge of the average LV direction to speed up their procedure. The localizer images were then segmented by thresholding and both the LV and RV were localized by comparison of the morphologic characteristics of the candidate objects. However, in our experience clinical image variability compromises the robustness of this method.

We relied on simple methods to automatically estimate heart location and orientation, in order to provide rapid feedback to higher level processes. The assumptions of our method are listed below. Any cases which violate these assumptions (eg congenital heart disease in which the LV and RV are transposed) would not be expected to be solved by our method.

1) The heart is the only large organ in the thorax with a spatial fundamental frequency equivalent to the cardiac cycle.
2) The orientation of the heart is similar across a wide variety of (non-congenital) cardiac diseases. (This assumption is validated below.)
3) The short axis (SA) slices are ECG gated and have been planned approximately orthogonal to the long axis of the LV. (We show that this assumption is not restrictive in practice.)
4) The positions of LV in adjacent slices are spatially and temporally coherent.
5) The septal myocardium is close to the centroid of the heart and has the LV and RV blood pools on each side. The boundary between the LV blood pool and septal myocardium is not degraded by large papillary muscles or trabeculations, which are typically not expected in this area anatomically.

The reminder of the paper is organized as follows. In Sect. 2, we describe the details of our method. In Sect. 3, we present the results from 330 patients recruited from 10 MR centers world-wide. Conclusion and discussion is provided in Sect. 4.

2 Method

Our method is based on the novel combination of the Fourier transform (FT) in the temporal domain with *a priori* orientation and shape information in space. The FT is employed to calculate an average image (DC) and a first harmonic (H1) magnitude image for each slice. Even in severely diseased hearts, this

method successfully identifies a large portion of the heart in most cases. The FT output is then used to derive a region of interest (ROI) and threshold which robustly delineates the LV. This four step process is summarized below:

1) Organize the frames slice by slice and apply the FT over time to obtain the DC and H1 images for each slice (Sect. 2.2).
2) Compute a ROI for each slice and the centroid of the whole heart from the H1 images (Sect. 2.3).
3) Find a pixel on the septal myocardium and compute the threshold level of the blood in the DC images (Sect. 2.4).
4) Threshold the DC images and locate the LV on all slices (Sect. 2.5).

2.1 Patient Data and Ground Truth

We examined 330 patients with cardiac and vascular disease enrolled in the CMR substudy to ONTARGET [9]. The patients had a range of disease histories: 294 had coronary artery disease, 46 had peripheral arterial disease, 111 had diabetes, 202 had hypertension and 192 had myocardial infarction with the total exceeding the number of patients due to multiple diagnoses. The patients were recruited in six countries and imaged using standard SSFP cardiac cine sequence on Siemens, Philips and GE scanners. Either prospectively or retrospectively gated cine images were acquired in six equally spaced SA locations from apex to base. Typical imaging parameters were TR / TE / flip / FOV = 30 ms / 1.6 ms / 60° / 360 mm, slice thickness 6 mm, image matrix 256 × 208. There were typically 25 temporal frames per slice, depending on heart rate. All cines were acquired during breath-hold of 8–15 seconds duration.

The ground truth for the heart location and orientation was determined manually by two experienced users operating independently on the end-diastolic phase images. The 3D orientation of the LV long axis V_x was defined by two points located in the middle of the LV blood pool at apex and base (Fig. 1a and 1b). The orientation of the RV V_y was defined by the centroid of the RV insertion points defined on each SA slice (Fig. 1c). The remaining axis V_z was oriented posterior to create a right handed coordinate system (Fig. 1d) [10]. The average directions $\overline{V}_x$, $\overline{V}_y$ and $\overline{V}_z$ of all cases were pre-computed for use in the automated method below.

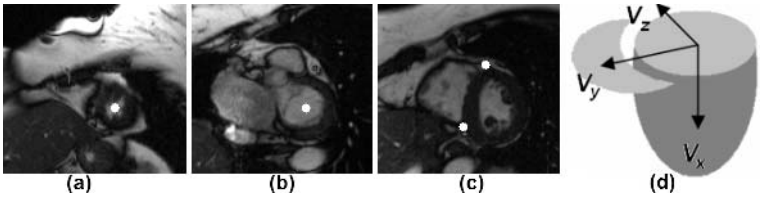


Fig. 1. Manual definition of heart orientation

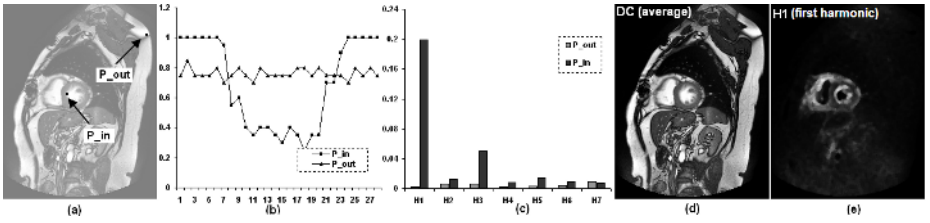


Fig. 2. Reasons for using FT over time sequence

2.2 Fourier Transform over Time

The heart is the only large structure in the thorax which has substantial motion with a frequency given by the heart rate, and this characteristic makes the heart distinguishable by analyzing the change in the intensity values over time at each pixel position. Fig. 2b shows typical pixel intensities through time. P_{in} is a pixel at the boundary between the LV blood pool and the septal myocardium. Its intensity value changes through a large range over time. P_{out} is also located close to the boundary of two different structures but is relatively static. Previously the standard deviation of the pixel values has been used to locate the heart [11, 12], however, we found that in about 20% of cases the standard deviation images were contaminated by excessive high frequency noise. We therefore computed the Fourier transform (Fig. 2c) and used the DC component (average image, Fig. 2d) and the H1 image (Fig. 2e) only. This method provides excellent delineation of the cardiac structures, as well as great vessels such as the aorta.

2.3 Fast ROI Analysis

A cardiac centroid and region of interest containing the heart were calculated from the H1 images for each slice as follows. Firstly, in order to reduce the effect of noise and signal from non-cardiac structures, the H1 images were filtered with a smoothing filter and all pixels with magnitude less than 5% of the maximum magnitude within the 3D volume were set to zero. Secondly, the ROI for each slice was iteratively refined. For each iteration, the centroid of the H1 image was computed for each slice. A 3D line was then fitted to the centroids of all slices by linear least squares. A distance distribution of all H1 pixels to the 3D line was calculated and weighted with pixel's intensity values. A Gaussian curve was fitted to the distribution and all pixels greater than a certain distance from the line were removed. The 3D centroid of the H1 volume was then computed and compared to the previous 3D centroid. The iteration terminated when the distance between successive 3D centroids was less than one pixel. In most cases, the iteration terminated after only one loop. Finally, the ROI was adjusted in each slice by using the Gaussian fitting method again but applied to the distribution of each slice separately. The result is shown in Fig. 3.

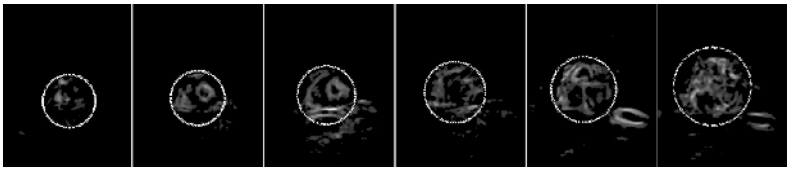


Fig. 3. Approximation of ROI in slices (apex to base from left to right)

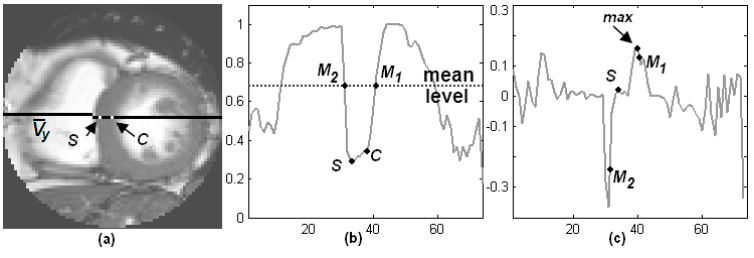


Fig. 4. Acquisition of threshold level

2.4 Parameters for Blood Pool Segmentation

In order to provide an initial segmentation of the LV blood pool, as well as a separation of RV and LV blood pools, we use the DC images within ROI (Fig. 4a) to locate a pixel within the septal myocardium and find the threshold level which best discriminates the blood and myocardial signals. Firstly, the mid-ventricular SA slice closest to the 3D centroid was chosen. The center of the ROI was obtained by intersecting the 3D central line (from Sect. 2.3) with the slice, marked C in Fig. 4. This point is almost always close to the interventricular septum. A line passing through C was defined in the direction of $\overline{V}_y$. The intensity of the DC image along this line (Fig. 4b) was then used to locate the septum by searching for a local minimum. Once a septal point S was found, the LV could be located on the $-\overline{V}_y$ side and the blood pool threshold could also be determined by searching for a local maximum on the gradient curve (Fig. 4c). To avoid the noise and uncertainty inherent in analyzing only a single sample, we also analyzed eight additional lines parallel to $\overline{V}_y$ and computed the average value of these boundary pixels.

2.5 LV Detection

The LV blood pool in the middle slice (i.e. the slice closest to the 3D centroid) was firstly localized on the $-\overline{V}_y$ side of S, as shown in Fig. 4a and Fig. 5a. A convex hull (Fig. 5b) was then used to reduce the impact of the papillary muscles, as in many other papers (e.g. [5, 8]).

To find the LV blood pools in adjacent slices, we modified the method proposed in [5]. The analysis was based on the binary images after thresholding, and assumed that the LV regions are spatially coherent between slices. The LV

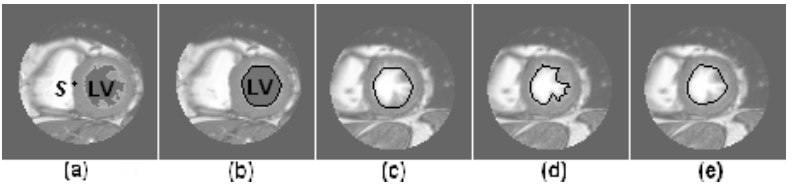


Fig. 5. Locating LV regions on the middle slice and its neighboring one

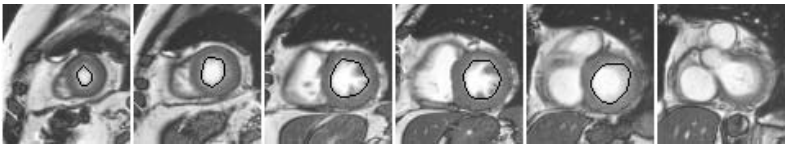


Fig. 6. Contours of the detected LV regions on slices

region detected in the middle slice was projected to its two neighboring slices and the binary objects on the neighboring slices were compared to the projected LV region. Rather than project in the direction normal to the slice [5], we projected the regions in the average long axis direction $\overline{V}_x$. This enables better robustness to the orientation of the image planes. After projection the most similar object in each slice was selected. The similarity of two objects was calculated by the area of the intersection divided by the area of union [5]. An example is shown in Fig. 5b-5e. Figure 5b is the middle slice with the detected LV region superimposed on it. The region is projected to its neighboring slice (Fig. 5c) and the most similar object is then found (Fig. 5d). The convex hull is finally applied to the new region (Fig. 5e).

With this method, the LV regions on all slices were located (Fig. 6). However, no region would be reported if similarity criteria was violated, as in the last slice of Fig. 6 . A 3D line was then fitted to the centroids of the LV regions.

3 Results

The fully automated method was implemented in Matlab and required approximately 4.13 seconds (not compiled) to run on a PC (Pentium IV 3.2GHz) for each case, excluding the DICOM file reading time.

Only one case out of 330 failed to detect the LV, in which the RV was found instead. The remaining 329 cases were used for quantitative comparison. The orientation error of the automated method was computed by the angle between the 3D central line (automated V_x) from the method and the ground truth V_x from the observers. This was compared to the orientation error of the manual planning in Table 1, which was obtained by the angle between the normal to the SA images and the ground truth V_x. The two errors are very similar in both observers and they are within twice the inter-observer error of 3.53 ± 2.43 between Observer A and B.

Table 1. Comparison of the orientation Errors in mean ± std format (unit: degree)

	V_x in automatic method	SA normal in Manual planning
V_x by Observer A	6.37 ± 4.41	6.30 ± 3.66
V_x by Observer B	6.14 ± 4.11	6.80 ± 3.95

Table 2. Distance errors between automated method and two observers (unit: mm)

	Apical Slice		Middle Slice		Basal Slice	
	V_y	V_z	V_y	V_z	V_y	V_z
V_x by Observer A	0.8 ± 3.0	-2.2 ± 3.7	1.2 ± 2.6	-2.6 ± 2.4	1.6 ± 4.3	-3.0 ± 3.7
V_x by Observer B	0.5 ± 2.9	-2.7 ± 3.3	0.7 ± 2.4	-3.3 ± 2.5	0.9 ± 3.9	-3.8 ± 4.0

We also computed the average manually determined orientations $\overline{V}_x$, $\overline{V}_y$ and $\overline{V}_z$ for each observer from the fiducial markers applied to the reference images. The inter-observer angular errors in $\overline{V}_x$, $\overline{V}_y$, $\overline{V}_z$ were 0.37, 3.01 and 3.03 degrees. The angular errors between $\overline{V}_x$ and the ground truth V_x were 10.56 ± 5.67 and 10.44 ± 5.50 for Observer A and Observer B. In 98% of cases, $\overline{V}_x$ was within 23.7 degrees of V_x, showing that the LV orientation is remarkably consistent across patients.

In order to compute the position errors, both the ground truth V_x and the automated V_x were intersected with the slices and the positions of the two intersection points were compared with each other in the ground truth coordinate system. The slices closest to the apex and base and the middle slice between these two end slices were used for the purposes of comparison. Table 2 shows the errors between the automated method and two observers in mean ± std. format. Figure 7 shows the distribution of the errors of the worst case (Observer A). The horizontal axis represents the V_y direction and the vertical is V_z, defined in Fig. 1d. It can be seen that the automatic results and the ground truth agree closely each other. There is a small bias in the V_z direction which may be caused by the differences between the manual and automatic measurement methods. The

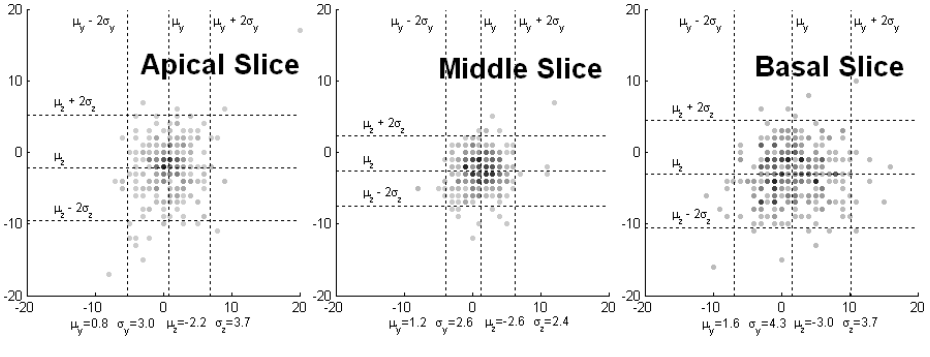


Fig. 7. Errors between ground truth V_x (Observer A) and automatic V_x on slices

ground truth V_x was measured only at end-diastole while the automatic V_x was based on the average images.

4 Conclusion and Discussion

The fully automatic method presented in this paper has been proven efficient and robust for detecting the location and orientation of the LV. The errors show that the automated method is as good as the manual scan planning results performed by experienced technicians. Our method assumes that cine images have been acquired approximately orthogonal to the LV long axis, and that average cardiac orientation axes have been pre-calculated using historical data, since this is within 23.7 degrees of the ground truth axes in 98% of cases. We also assume that the images are ECG gated; however, the method could be extended to self gated acquisitions by automatic detection of the heart rate.

References

1. Santarelli, M., Positano, V., Michelassi, C., Lombardi, M., Landini, L.: Automated cardiac MR image segmentation: theory and measurement evaluation. *Med. Eng. Phys.* **25** (2003) 149–159
2. Lorenzo-Valdés, M., Sanchez-Ortiz, G., Mohiaddin, R., Rueckert, D.: Segmentation of 4D cardiac MR images using a probabilistic atlas and the EM algorithm. In: *MICCAI'03*. (2003) 440–450
3. 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 Trans. Med. Imag.* **21**(9) (2002) 1167–1178
4. Kaus, M.R., von Berg, J., Niessen, W., Pekar, V.: Automated segmentation of the left ventricle in cardiac MRI. In: *MICCAI'03*. (2003) 432–439
5. Montillo, A., Metaxas, D., Axel, L.: Automated segmentation of the left and right ventricles in 4D cardiac SPAMM images. In: *MICCAI'02*. (2002) 620–633
6. Spreeuwers, L., Breeuwer, M.: Automatic detection of the myocardial boundaries of the right and left ventricle. *SPIE: Med. Imag.* **4322** (2001) 1207–1217
7. Danilouchkine, M., Westenberg, J., Reiber, J., Lelieveldt, B.: Accuracy of short-axis cardiac MRI automatically derived from scout acquisitions in free-breathing and breath-holding modes. *MAGMA* **18** (2005) 7–18
8. Jackson, C., Robson, M., Francis, J., Noble, J.: Automatic planning of the acquisition of cardiac MR images. In: *MICCAI'03*. (2003) 541–548
9. Anderson, C.: Rationale and design of the cardiac magnetic resonance imaging substudy of the ONTARGET trial programme. *J. Int. Med. Res.* **33**(4) (2005) 50A–57A
10. Young, A., Cowan, B., Thrupp, S., Hedley, W., Dell'Italia, L.: Left ventricular mass and volume: fast calculation with guide-point modeling on MR images. *Radiology* **216** (2000) 597–602
11. Sörgel, W., Vaerman, V.: Automatic heart localization from 4D MRI datasets. *SPIE: Med. Imag.* **3034** (1997) 333–344
12. Gering, D.: Automatic segmentation of cardiac MRI. In: *MICCAI'03*. (2003) 524–532

Pairwise Active Appearance Model and Its Application to Echocardiography Tracking

S. Kevin Zhou[1], Jie Shao[2,⋆], Bogdan Georgescu[1], and Dorin Comaniciu[1]

[1] Integrated Data Systems, Siemens Corporate Research, Inc., Princeton, NJ, USA
{shaohua.zhou, bogdan.georgescu, dorin.comaniciu}@siemens.com
[2] Center for Automation Research, University of Maryland, College Park, MD, USA
shaojie@cfar.umd.edu

Abstract. We propose a pairwise active appearance model (PAAM) to characterize statistical regularities in shape, appearance, and motion presented by a target that undergoes a series of motion phases, such as the left ventricle in echocardiography. The PAAM depicts the transition in motion phase through a Markov chain and the transition in both shape and appearance through a conditional Gaussian distribution. We learn from a database the joint Gaussian distribution of the shapes and appearances belonging to two consecutive motion phases (i.e., a pair of motion phases), from which we analytically compute the conditional Gaussian distribution. We utilize the PAAM in tracking the left ventricle contour in echocardiography and obtain improved tracking results in terms of localization accuracy when compared with expert-specified contours.

1 Introduction

Characterizing shape, appearance, and motion is an important research topic in medical imaging applications. There exists a wide literature on this topic; we here only focus on one particular type of approach – active models. Active shape model (ASM) [1] depicts shape statistics using principal component analysis (PCA). Active appearance model (AAM) [2] extends the ASM to model the appearance too with both shape and appearance are jointly modeled by PCA. The ASM and AAM are applicable to images only. To deal with a video, active appearance motion model (AAMM) [3] extends the AAM to characterize the motion in the video and is used for segmenting a spatiotemporal object. One restriction of the AAMM is that no global motion is allowed before neighboring frames; hence the AAMM is not applicable to online tracking. Attempting to solving the tracking task, we present a novel model called pairwise active appearance model (PAAM) that characterizes shape, appearance and motion in one treatment.

We apply the PAAM for tracking the left ventricle in 2D cardiac ultrasonography (or echocardiography). Echocardiography tracking [4, 5, 6, 7, 8, 9, 10] is challenging due to severe imaging artifacts. Artifacts arise from ultrasound speckle

⋆ The work was done when Shao was with SCR in summer 2005. We thank Dr. S. Krishnan of Siemens Medical Solutions for providing the data.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 736–743, 2006.

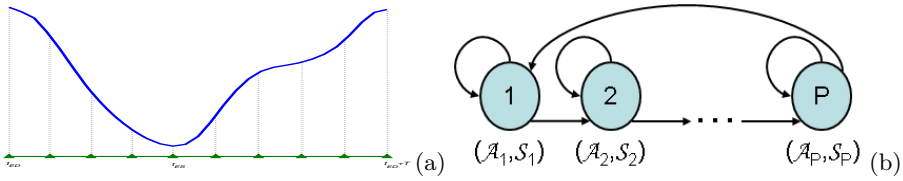


Fig. 1. (a)A cardiac cycle is divided into $P = 9$ motion phases. The blue curve shows the LV volume. (b) A pictorial illustration of the PAAM.

noise (due to signal refraction and reverberation), signal dropout, and are also characterized by missing, fake, and improperly located anatomic structures. The left ventricle (LV) appearance changes are caused by fast movement of heart muscle, respiratory inferences, unnecessary transducer movement, etc. With the aid of the PAAM, we successfully regularize the optical flow measurement and obtain improved shape tracking results.

2 Pairwise Active Appearance Model

Assuming that the target of interest undergoes a series of P motion phases indexed by $p = \{1, 2, ..., P\}$. Fig. 1(a) shows an example of dividing a cardiac cycle into four equally space motion phases in systole and five motion phases in diastole.

Fig. 1(b) illustrates the underlying principle of the PAAM. (i) The PAAM depicts the transition in motion phase through a Markov chain: it either stays at the current motion phase or proceeds to the next one. For example, in the cardiac example, given the end of diastole (ED) and the end of systole (ES) frames, one can easily determine which motion phase the current frame belongs to. (ii) The PAAM depicts the transition in both shape and appearance through a conditional Gaussian distribution. We learn from a database the joint Gaussian distribution of the shapes and appearances belonging to two consecutive motion phases (i.e., a pair of motion phases), from which we analytically compute the conditional Gaussian distribution.

2.1 Learning the PAAM

The shape is represented by M_s landmark points, or equivalently a $2M_s$-dimensional vector $\mathcal{S}$. The appearance $\mathcal{A}$ is represented by an M_g-dimensional vector. We concatenate the shape and appearance vectors at two consecutive motion phases to form paired data: $\mathbf{s}_p = [\mathcal{S}_p^{\mathsf{T}} \mid \mathcal{S}_{p-1}^{\mathsf{T}}]^{\mathsf{T}}$ and $\mathbf{a}_p = [\mathcal{A}_p^{\mathsf{T}} \mid \mathcal{A}_{p-1}^{\mathsf{T}}]^{\mathsf{T}}$, where $p \in \{1, 2, \ldots, P\}$ is the phase index. We assume that $\mathcal{S}_0 \doteq \mathcal{S}_P$ and $\mathcal{A}_0 \doteq \mathcal{A}_P$.

We follow the procedure of learning the AAM for each pair of motion phases: (i) Construct the shape subspace based on $\mathbf{s}_p$ using the principal component analysis (PCA). The subspace can be represented by:

$$\mathbf{s}_p \approx \bar{\mathbf{s}}_p + \mathbf{P}_p^{<s>}\mathbf{b}_p^{<s>}, \tag{1}$$

where $\mathbf{P}^{<s>}$ is a subspace matrix (eigenvectors) describing a sufficient fraction of the total shape variation, $\mathbf{b}^{<s>}$ is a vector containing the combination coefficients for each of the eigenvectors. (ii) Similarly, construct the appearance subspace based on $\mathbf{a}_p$ using the PCA.

$$\mathbf{a}_p \approx \bar{\mathbf{a}}_p + \mathbf{P}_p^{<a>}\mathbf{b}_p^{<a>}. \tag{2}$$

(iii) Apply a third PCA to the combination of shape and appearance:

$$\mathbf{b}_p = \begin{bmatrix} \mathbf{b}_p^{<s>} \\ \mathbf{W}_p^{<a>}\mathbf{b}_p^{<a>} \end{bmatrix} \approx \mathbf{Q}_p\mathbf{c}_p = \begin{bmatrix} \mathbf{Q}_p^{<s>} \\ \mathbf{Q}_p^{<a>} \end{bmatrix} \mathbf{c}_p, \tag{3}$$

where $\mathbf{W}_p^{<a>}$ is a diagonal matrix that balances the energy discrepancy between the shape and appearance models, $\mathbf{Q}_p$ is the eigenvector matrix, and $\mathbf{c}_p$ is a latent vector that controls both the shape and appearance models.

We recapitulate the PAAM in a statistical jargon. Denote both shape and appearance by $\mathbf{z} = [\mathcal{S}^\mathsf{T}, \mathcal{A}^\mathsf{T}]^\mathsf{T}$. For the p^{th} pair of motion phases, its distribution $p(\mathbf{z}_p, \mathbf{z}_{p-1}) = p(\mathcal{S}_p, \mathcal{A}_p, \mathcal{S}_{p-1}, \mathcal{A}_{p-1})$ is Gaussian, whose mean and covariance matrix are expressed as:

$$\mu_p = \begin{bmatrix} \mu_p^{<z>} \\ \mu_{p-1}^{<z>} \end{bmatrix}, \quad \mathbf{\Sigma}_p = \begin{bmatrix} \mathbf{\Sigma}_{p,p}^{<z>} & \mathbf{\Sigma}_{p,p-1}^{<z>} \\ \mathbf{\Sigma}_{p-1,p}^{<z>} & \mathbf{\Sigma}_{p-1,p-1}^{<z>} \end{bmatrix}.$$

It is easy to see that the conditional probability $p(\mathbf{z}_p|\mathbf{z}_{p-1})$, which is actually used in tracking, is also Gaussian with mean and covariance matrix given as:

$$\mu_{p|p-1}^{<z>} = \mu_p^{<z>} + \mathbf{\Sigma}_{p,p-1}^{<z>}[\mathbf{\Sigma}_{p-1,p-1}^{<z>}]^{-1}(\mathbf{z}_{p-1} - \mu_{p-1}^{<z>}), \tag{4}$$

$$\mathbf{\Sigma}_{p|p-1}^{<z>} = \mathbf{\Sigma}_{p,p}^{<z>} - \mathbf{\Sigma}_{p,p-1}^{<z>}[\mathbf{\Sigma}_{p-1,p-1}^{<z>}]^{-1}\mathbf{\Sigma}_{p-1,p}^{<z>}. \tag{5}$$

In practice, when the Gaussian assumption is not satisfactory, we group the data into several clusters and learn the PAAM for each cluster to handle possible data nonlinearity.

2.2 Using the PAAM in Tracking

Tracking algorithms can be broadly divided into two categories, depending on the way in which online observations and offline learned models are integrated. (i) The models are embedded into the so-called observation likelihood. The motion parameters are used to deform the observation to best fit the likelihood. An example is the famous active appearance model (AAM) [2]. (ii) Generic optical flow computation is first conducted for each landmark; learned models are then applied to regularize the overall shape. An example is the work of Zhou *et al.* [10], which is referred to as fusion approach. We follow [10] due to its flexibility. The fusion approach consists of two processes: *observation* and *fusion*. The observation process computes optical flow for individual landmarks and the fusion process regularizes the whole contour. In this paper, we mainly focus on the fusion process. In the observation process, we utilize our earlier approach

[9] where a *nonparametric local appearance model* (NLAM) is constructed on the fly to model the shape and appearance at a point level. The output of the observation process is the location and covariance matrix of the landmarks as well as the appearance and its uncertainty.

At time instant t, the fusion process derives an optimal solution $\mathbf{z}_t^*$ that minimizes the fusion cost $d^2_{t|t-1} = d^2_{t|t-1,1} + d^2_{t|t-1,2}$, where

$$d^2_{t|t-1,i} = (\mathbf{z}_t - \mathbf{z}_{t|t-1,i})^{\mathsf{T}} \mathbb{C}^{-1}_{t|t-1,i} (\mathbf{z}_t - \mathbf{z}_{t|t-1,i}); \quad i = 1, 2, \tag{6}$$

and $\mathbf{z}_{t|t-1,i}$ and $\mathbb{C}_{t|t-1,i}$ are the mean vector and covariance matrix, respectively. The first distance $d^2_{t|t-1,1}$ in (6) arises from the observation process that provides the mean vector $\mathbf{z}_{t|t-1,1}$ and the covariance matrix $\mathbb{C}_{t|t-1,1}$. The second distance $d^2_{t|t-1,2}$ in (6) is from the PAAM (refer to (4) and (5)). There are two possible situations from time $t-1$ to t: (a) there is no transition in the motion phase, i.e., staying at the same motion phase p; or (b) there is a transition in the motion phase from $p-1$ to p.

$$\mathbf{z}_{t|t-1,2} = \mu_p^{<z>}, \quad \mathbb{C}_{t|t-1,2} = \mathbf{\Sigma}_{p,p}^{<z>}; \text{ if (a).} \tag{7}$$

$$\mathbf{z}_{t|t-1,2} = \mu_{p|p-1}^{<z>}, \quad \mathbb{C}_{t|t-1,2} = \mathbf{\Sigma}_{p|p-1}^{<z>}; \text{ if (b).} \tag{8}$$

When evaluating the above $\mu_{p|p-1}^{<z>}$ exactly defined in (4), we use $\mathbf{z}_{p-1} = \mathbf{z}_{t-1}^*$. It is easy to determine (a) or (b) in echocardiography by using the cardiac period T, the ED frame t_{ED}, and the ES frame t_{ES}. All these information is directly available from the video sequence file.

We observe that when a motion transition happens, using the conditional probability $p(\mathbf{z}_p|\mathbf{z}_{p-1})$ is beneficial because $\mathbf{z}_{t|t-1,2}$ is always updated during the iterations and hence adaptive to the previous observation $\mathbf{z}_{t-1}^*$. On the other hand, the covariance matrix $\mathbf{\Sigma}_{p|p-1}^{<z>}$ is fixed and hence pre-computable during training, which improves computational efficiency.

Usually, $\mathbb{C}_{t|t-1,2}$ is singular due to the high dimensionality of the shape and appearance vectors, thereby leading to a *non-orthogonal subspace projection* problem. Suppose the rank of $\mathbb{C}_{t|t-1,2}$ is q and its rank-q SVD is $\mathbb{C}_{t|t-1,2} = \mathtt{U}_q \Lambda_q \mathtt{U}_q^{\mathsf{T}}$, the best fusion estimator that minimizes the fusion cost $d^2_{t|t-1}$ is the so-called best linear unbiased estimate [10]:

$$\mathbf{z}_t^* = \mathtt{U}_q (\mathtt{U}_q^{\mathsf{T}} \mathbb{C}^{-1}_{t|t-1,1} \mathtt{U}_q + \Lambda_q^{-1})^{-1} (\mathtt{U}_q^{\mathsf{T}} \mathbb{C}^{-1}_{t|t-1,1} \mathbf{z}_{t|t-1,1} + \Lambda_q^{-1} \mathtt{U}_q^{\mathsf{T}} \mathbf{z}_{t|t-1,2}). \tag{9}$$

In practice, because we cluster the data and learn several sub-models for each pair of motion phases, the sub-model with the smallest fusion cost is selected.

3 Experimental Results

We have 400 A4C (apical four-chamber) sequences and 320 A2C (apical two-chamber) sequences. In total, there are about 11000 A4C frames and about 9200 A2C frames. We used 5-fold cross validation for performance evaluation. The

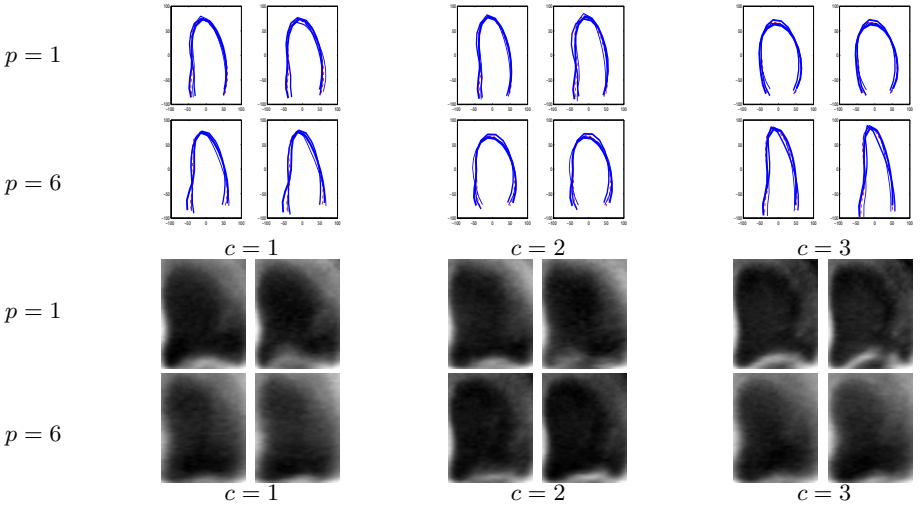


Fig. 2. Example of shape and appearance subspaces of the trained PAAM. In our experiments we trained three sub-models for each pair of motion phases. c: cluster index, p: phase index. Rows in the figure correspond to clusters; columns correspond to phases. In the shape model, the red dot lines represent the means of the subspaces, while the three blue solid lines in each plot represent three eigenvectors associated with the top three eigenvalues in the corresponding subspaces.

ground truth contours were generated by experts. During testing, we assumed manual initialization at the middle frame between the ED and ES frames.

Preprocessing. Before training, we performed the following preprocessing steps: (a) Video frames are sampled and classified to different phases. Global appearance patches are cropped out from each sampled frame and then rigidly aligned to a mean shape in a 50×40 template using the thin-plate splines warping algorithm. (b) Since echocardiograms have highly non-Gaussian intensity histograms, we applied a nonlinear ultrasound-specific normalization method [3] to transform the non-Gaussian intensity histogram to have a normal distribution. However, since this is only for the appearance, the joint space of shape and appearance is hardly Gaussian even after this transformation. (c) The shape consists of 17 control points, which means that the dimensionality of the shape vector is 34. The appearance patch contains $50 \times 40 = 2000$ pixels. Since such a high dimension requires expensive computation, we applied a preprocessing PCA to reduce the dimensionality of the appearance from 2000 to around 1000, before feeding them to train the PAAM. Using the preprocessed data, we trained the PAAM with $P = 9$ components, each component containing three sub-models. Fig. 2 illustrates the learned shape and appearance subspaces.

Two contour distances. To evaluate the tracking performance, we need to measure the proximity between two contours. Rather than using the rigid Euclidean distance to measure the distance between two landmark points, we propose a *segmental Hausdorff distance* (segHD) that allows certain degree of non-rigidity.

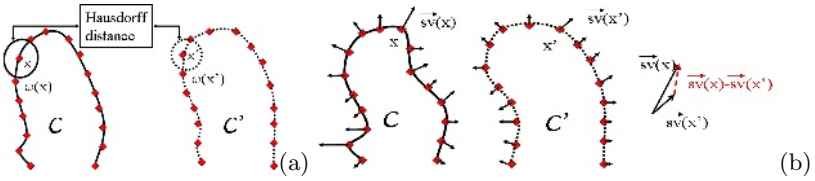


Fig. 3. (a) Segmental Hausdorff distance. (b) Surprisal vector distance.

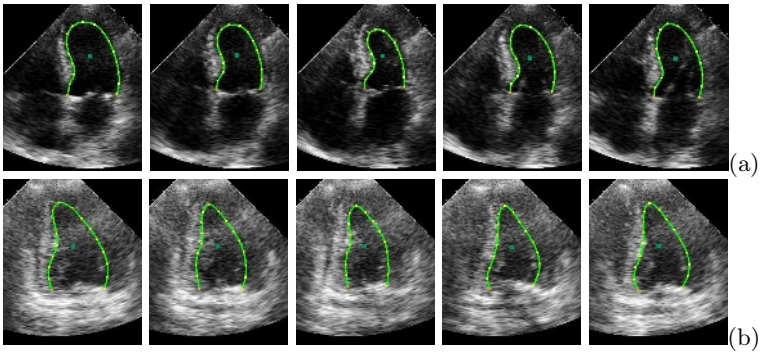


Fig. 4. The snapshots of the tracking results of (a) an A4C sequence and (b) an A2C sequence

As illustrated in Fig. 3(a), the segHD between two corresponding landmark points $\mathtt{x}$ and $\mathtt{x}'$ on the two curves $\mathcal{C}$ and $\mathcal{C}'$, respectively, is defined as the Hausdorff distance (HD) between two segments $\omega(\mathtt{x})$ and $\omega(\mathtt{x}')$, where $\omega(\mathtt{x})$ defines a segment around $\mathtt{x}$ on the curve $\mathcal{C}$. We further take the mean of the segHD of all landmarks as the distance between $\mathcal{C}$ and $\mathcal{C}'$, denoted by $d_{segHD}(\mathcal{C}, \mathcal{C}')$.

$$shd(\mathtt{x}, \mathtt{x}') = HD(\omega(\mathtt{x}), \omega(\mathtt{x}')); d_{segHD}(\mathcal{C}, \mathcal{C}') = \{\int_{\mathtt{x}} shd(\mathtt{x}, \mathtt{x}')d\mathcal{C}\}/\{\int_{\mathtt{x}} d\mathcal{C}\}. \quad (10)$$

The segHD measures only the 'physical distance' between two contours, ignoring their curvedness. Even when the two contours $\mathcal{C}'$ and $\mathcal{C}''$ have the same distance to the ground truth contour $\mathcal{C}$ in terms of d_{segHD}, $\mathcal{C}'$ and $\mathcal{C}''$ can be differently perceived because they present different curvedness. Feldman and Singh proposed [11] a *surprisal vector* $\overrightarrow{sv}$ to quantify how the curve is perceived. Fig. 3(b) illustrates the surprisal vector. The direction of $\overrightarrow{sv}$ is the same as the outward normal direction and the magnitude $|\overrightarrow{sv}|$ is a function of curvature. When at the highly-curved part of the contour, the $|\overrightarrow{sv}|$ is large; when at the flat part, it is small. Using the surprisal vector, we compute a *surprisal vector distance* $d_{surp}(\mathcal{C}, \mathcal{C}')$ to characterize the proximity of two contours in their curvedness.

$$surp(\mathtt{x}, \mathtt{x}') = ||\overrightarrow{sv}(\mathtt{x}) - \overrightarrow{sv}(\mathtt{x}')||^2; d_{surp}(\mathcal{C}, \mathcal{C}') = \{\int_{\mathtt{x}} surp(\mathtt{x}, \mathtt{x}')d\mathcal{C}\}/\{\int_{\mathtt{x}} d\mathcal{C}\}. \quad (11)$$

Tracking performance. We first compared four methods whose results are reported in Table 1 using the median and standard deviation of the contour dis-

Table 1. Tracking performance based on (a,b) the segmental Hausdorff distance and (c) the surprisal vector distance using (a,c) the SSD, CD_2, NLAM, and PAAM methods and (b)the ASM, AAM, PASM, and PAAM methods

(a) Sequences	Segmental Hausdorff distance d_{segHD} (pixels)			
	SSD	CD_2	NLAM	PAAM
A2C	10.8612 ± 2.2621	7.9392 ± 1.5645	2.7042 ± 0.6732	2.6275 ± 0.6623
A4C	11.0310 ± 2.5927	7.3640 ± 2.3561	2.5291 ± 0.6076	2.4588 ± 0.5550
(b) Sequences	Segmental Hausdorff distance d_{segHD} (pixels)			
	ASM	AAM	PASM	PAAM
A2C	2.6901 ± 0.6611	2.6844 ± 0.6881	2.6849 ± 0.6951	2.6275 ± 0.6623
A4C	2.5191 ± 0.5915	2.4776 ± 0.5614	2.5059 ± 0.5930	2.4588 ± 0.5550
(c) Sequences	Surprisal vector distance d_{surp}			
	SSD	CD_2	NLAM	PAAM
A2C	0.3204 ± 0.1256	0.0957 ± 0.1197	0.0352 ± 0.0514	0.0098 ± 0.0110
A4C	0.3024 ± 0.1147	0.0995 ± 0.1006	0.0345 ± 0.0586	0.0096 ± 0.0097

tances for all testing video sequences in five folds. The SSD means the general optical flow method using the sum of squared distance similarity function; the CD_2 using the similarity function in [8], which considers a simplified ultrasound image formation; and the NLAM using the method in [9]. The PAAM means regularizing the NLAM results using the PAAM. From Table 1(a), we observe that the NLAM improves the tracking results significantly in terms of the segHD, compared with SSD and CD_2. Using the PAAM further decreases the segHD by some margin. The advantage of using the PAAM is highlighted when the surprisal vector distance is used. Using the NLAM only often results in a wiggly contour as every landmark is tracked independently. However, the PAAM successfully regularizes the wiggly contour into a smooth one. This regularization is quantized by the surprisal vector distance: the PAAM yields significant lower error as reflected in Table 1(c). Fig. 4 shows the tracking contours overlaid on sample frames of an A4C sequence and an A2C sequence.

Next, we show that the effectiveness of shape, appearance and motion information when used as prior knowledge. Table 1(b) shows the performance after regularizing the NLAM results using four different prior models in the fusion process. The ASM means using the *phase-separate* active shape model (ASM) only, without the pairwise model. In other words, we trained ASMs for each of the nine phases. No motion and appearance information is interpreted in the model. The AAM model means using the *phase-separate* AAM only, which takes into account shape and appearance. The third model uses the PASM (pairwise ASM) model, with shape and motion but no appearance information involved. The last model is the PAAM model that jointly considers shape, appearance and motion. Table 1(b) tells that using more prior information results in decreased tracking error. It also indicates the order of the importance of the three elements: *shape* > *appearance* > *motion*. For example, the fact that the AAM provides better performance than the PASM suggests that the appearance information contributes more to the entire system than the motion information.

Comparison with the AAMM. We summarize the main differences between the AAMM and PAAM since both are able to capture shape, appearance, and motion. First, the AAMM is suitable to segment a spatiotemporal target

in a sequence, but hardly fits to an online tracking task. Second, the AAMM assumes that the motion only comes from the heart beating. Little or no motion is introduced by external factors such as ultrasound transducer movement that is always present in practice. Third, the AAMM lacks adaptability to different cases since it falls in the 'observation explains model' category. Finally, the AAMM is very high-dimensional, causing ineffective modeling capability due to difficulty in collecting enough data to cover desired variations, and expensive computations in both training and testing. The proposed PAAM contains the promising properties of the AAMM, with flexibility and adaptivity integrated. It also enhances the modeling capability and computational efficiency.

4 Conclusion

We have proposed the PAMM to represent shape, appearance, and motion information. The shape and appearance knowledge is described by the model subspaces, while the inter-phase motion is described by paired data. We integrated the model into a fusion algorithm for tracking. In the experiments, we applied the tracker to a large study of LV tracking and demonstrated robustness and accuracy, using the segmental Hausdorff distance and surprisal vector distance, in tracking both A4C and A2C echocardiographic sequences.

References

1. Cootes, T., Taylor, C.: Active shape models - 'smart snakes'. In: BMVC. (1992)
2. Cootes, T., Edwards, G., Taylor, C.: Active appearance models. PAMI **23** (2001) 681–685
3. Bosch, J., et al.: Automatic segmentation of echocardiographic sequences by active appearance motion models. IEEE Trans. Medical Imaging **21** (2002) 1374–1383
4. Mikic, I., Krucinski, S., Thomas, J.: Segmentation and tracking in echocardiographic sequences: Active contours guided by optical flow estimates. IEEE Trans. Medical Imaging **17** (1998) 274–284
5. Jacob, G., Noble, A., Blake, A.: Robust contour tracking in echocardiographic sequence. In: Proc. Intl. Conf. on Computer Vision. (1998) 408–413
6. Ledesma-Carbayo, M., et al.: Cardiac motion analysis from ultrasound sequences using non-rigid registration. In: MICCAI. (2001) 889–896
7. Jacob, G., et al.: A shape-space-based appraoch to tracking myocardial borders and quantifying regional left-ventricular function applied in echocardiography. IEEE Trans. Medical Imaging **21** (2002) 226–238
8. Boukerroui, D., Alison, J., Brady, M.: Velocity estimation in ultrasound images: A block matching approach. In: IPMI. (2003) 586–598
9. Georgescu, B., Zhou, X., Comaniciu, D., Rao, B.: Real-time multi-model tracking of myocardium in echocardiography using robust information fusion. In: MICCAI. (2004) 777–785
10. Zhou, X.S., Comaniciu, D., Gupta, A.: An information fusion framework for robust shape tracking. PAMI **27** (2005) 115–129
11. Feldman, J., Singh, M.: Information along contours and object boundaries. Psychological Review **112** (2005) 243–252

Cardiac Motion Recovery: Continuous Dynamics, Discrete Measurements, and Optimal Estimation

Shan Tong[1] and Pengcheng Shi[1,2]

[1] Department of Electronic and Computer Engineering,
Hong Kong University of Science and Technology, Hong Kong
[2] School of Biomedical Engineering, Southern Medical University, China*

Abstract. A sampled-data filtering framework is presented for cardiac motion recovery from periodic medical image sequences. Cardiac dynamics is a continuously evolving physiological process, whereas the imaging data can provide only sampled measurements at discrete time instants. Stochastic multi-frame filtering frameworks are constructed to couple the continuous cardiac dynamics with the discrete measurements, and to deal with the parameter uncertainty of the biomechanical constraining model and the noisy nature of the imaging data in a coordinated fashion. The state estimates are predicted according to the continuous-time state equation between observation time points, and then updated with the new measurements obtained at discrete time instants, yielding physically more meaningful and more accurate estimation results. Both continuous-discrete Kalman filter and sampled-data H_∞ filter are applied, and the H_∞ scheme can give robust estimation results when the noise statistics is not available a priori. The sampled-data estimation strategies are validated through synthetic data experiments to illustrate their advantages and on canine MR phase contrast images to show their clinical relevance.

1 Introduction

Myocardial kinematics reveals critical information about cardiac physiology and pathology, and accordingly there have been abundant efforts devoted to cardiac motion recovery from medical image sequences [1]. Cardiac dynamics is a physical/physiological process which evolves continuously in time, consisting of electrical propagation, electromechanical coupling, and biomechanics-based tissue deformation. However, the imaging/imaging-derived measurements can provide only sampled measurements at discrete time instants, such as computed tomography (CT) images, magnetic resonance (MR) images and ultrasound (US) images. In consequence, the kinematics estimation should be performed on a hybrid system with *continuous-time dynamics* and *discrete-time measurements*.

* Thanks to IBM PhD fellowship for supporting Shan Tong. This work is supported in part by China National Basic Research Program(973-2003CB716100), National Natural Science Foundation of China(60403040) and by Hong Kong Research Grants Council(CERG-HKUST6151/03E).

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 744–751, 2006.

Incorporating uncertainties is also essential in cardiac image analysis. As additional constraining models are needed for the ill-posed motion recovery problem, system uncertainties should be incorporated into the constraining models to obtain the flexibility of dealing with subject-dependent data sets. On the other hand, due to the noisy nature of the imaging/imaging-derived measurements, data uncertainties also need to be properly considered.

Formulating the cardiac image analysis problem in a hybrid system, our task is to construct an estimation framework so as to couple the continuous cardiac dynamics with the discrete imaging-derived measurements, and to deal with the system uncertainty of a priori constraining models and the data uncertainty of a posteriori noisy measurements in a coordinated effort. Multi-frame filtering strategies have been applied to tackle this problem. In previous works [2, 3], the system dynamics is first converted to discrete time to perform the analysis, but such conversion involves several assumptions and approximations, which are unrealistic in most real cases, especially for large sampling interval T.

In this paper, we propose a sampled-data filtering framework for kinematics estimation of the hybrid cardiac analysis system. *Instead of discretizing the system dynamics, state estimates are predicted according to the original continuous state equation and then updated with new measurements at discrete observation time points, which is physically more meaningful for the continuously evolving cardiac dynamics. As approximation errors in the discretization are avoided, estimates of higher accuracy are obtained.* Both minimum-mean-squared-error (MMSE) and mini-max H_∞ optimization criteria are adopted to couple the continuous dynamics with the discrete observations, and accordingly continuous-discrete Kalman filter and sampled-data H_∞ filter are applied for motion recovery. Synthetic experiments have been conducted to validate the advantages of sampled-data filtering and the robustness of H_∞ strategy. Results from canine MR images are also presented to show the clinical relevance of this framework.

2 Methodology

2.1 Continuous Biomechanics-Based Myocardial Dynamics

In biomechanics-based motion analysis, the structure and material of the myocardium should be properly modeled. As it has been shown that the material properties along and cross the myofibers are substantially different [4], we adopt the linear anisotropic material model for a realistic nonrigid motion recovery.

Under spatio-temporal biomechanical constrains, the dynamic equation for myocardium is derived with the principle of minimum potential energy [2]:

$$M\ddot{U} + C\dot{U} + KU = R \tag{1}$$

where M, C and K are the mass, damping and stiffness matrices respectively, R is the load vector, and U is the displacement vector. K is related to the material-specific Young's modulus, Poisson's ratio, and fiber orientations [4]. Rayleigh damping is assumed with $C = \alpha M + \beta K$.

To apply filtering strategies to our estimation problem, the dynamic equation (1) is transformed into a state-space representation of a *continuous-time* system:

$$\dot{x}(t) = A(t)x(t) + B(t)w(t) + \tilde{v}(t) \tag{2}$$

where the state vector $x(t)$, the input vector $w(t)$, the system matrix $A(t)$, and the input gain $B(t)$ are as follows:

$$x(t) = [U(t), \dot{U}(t)]', w(t) = [0, R(t)]', A(t) = \begin{bmatrix} 0 & I \\ -M^{-1}K & -M^{-1}C \end{bmatrix}, B(t) = \begin{bmatrix} 0 & 0 \\ 0 & M^{-1} \end{bmatrix},$$

and $\tilde{v}(t)$ is the *process noise* describing the disturbances/uncertainties in cardiac dynamics which come from the external load and the subject-dependent material properties. As the material model parameters and the geometry of heart will change over time, this state-space equation represents a time-varying system.

2.2 Discrete Measurements from Medical Imaging Data

Although cardiac dynamics has a continuously evolving nature, medical imaging data can provide available measurements $y(k)$ only at discrete observation time instants, which are expressed in a *discrete-time* measurement equation:

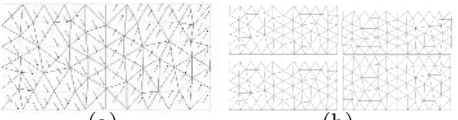
$$y(k) = Dx(k) + e(k) \tag{3}$$

where D is the measurement matrix, $e(k)$ is the *discrete-time measurement noise* accounting for the imaging data uncertainties, and k is used to denote the observation time instant $t = kT$. D describes the relationship of the sampled system states $x(k)$ and the discrete imaging data $y(k)$, and many different kinds of data can be incorporated by proper choices of D. In our case, the imaging data provides the displacement and velocity information, which can be obtained from MR tagging data and MR phase contrast images respectively.

2.3 Optimal Motion Estimation: A Sampled-Data Filtering Framework

From the above analysis, cardiac image analysis should be performed on a hybrid system with *continuous-time* dynamics and *discrete-time* measurements, and our goal is to construct a filtering framework so as to estimate the states of the continuous dynamics given the sampled measurements in some optimal senses.

In order to obtain accurate and physiologically meaningful estimates for the continuous cardiac dynamics, we propose a *sampled-data filtering* framework for kinematics estimation in such a hybrid system described by equations (2) and (3). In contrast to the system discretization approach in [2, 3] (a detailed comparison is presented in 2.4), the estimate in the sampled-data filtering framework is given directly in terms of the continuous dynamics and the sampled measurements. Without discretizing equation (2), the state estimates are propagated according to the original continuous-time state equation and then updated with new measurements at discrete observation time instants. Both MMSE and minimax H_∞ optimization criteria are adopted, and accordingly continuous-discrete Kalman filter and sampled-data H_∞ filter are applied for the motion recovery.



(a) (b)

Fig. 1. (a): Material composition with the arrows denoting fiber orientations. (b): Deformed geometry at selected frames #4, #8, #12, #16.

Continuous-Discrete Kalman Filtering. Assuming process and measurement noises $\tilde{v}(t)$ and $e(k)$ to be Gaussian distributed, zero-mean and white $(E[\tilde{v}(t)] = 0, E[\tilde{v}(t)\tilde{v}(\tau)'] = \tilde{Q}(t)\delta(t-\tau), E[e(k)] = 0, E[e(k)e(j)'] = V(k)\delta_{kj})$, continuous-discrete Kalman filter yields the optimal estimate in the MMSE sense.

Continuous-discrete Kalman filter is a recursive filter consisting of a prediction stage and an update stage [5]. In the prediction stage, i.e. between two observation time instants, the state estimate $\hat{x}(t)$ and its covariance $P(t)$ propagate according to the following two prediction equations [5]:

$$\dot{\hat{x}}(t) = A(t)\hat{x}(t) + B(t)w(t) \tag{4}$$

$$\dot{P}(t) = A(t)P(t) + P(t)A(t)' + \tilde{Q}(t) \tag{5}$$

Define $\hat{x}(kT^-) = \lim_{\varepsilon \to 0} \hat{x}(kT - \varepsilon)$. The predicted state $\hat{x}(kT^-)$ and its associated covariance $P(kT^-)$ can be obtained by solving equations (4) and (5) on the time interval $[(k-1)T, kT]$, with the previous state estimate $\hat{x}((k-1)T)$ and covariance $P((k-1)T)$ as the initial conditions of the differential equations.

In the update stage, i.e. at the observation time $t = kT$ with new measurements available, the state estimate and its covariance are updated by the standard discrete Kalman filter as follows:

$$\hat{x}(kT) = \hat{x}(kT^-) + W(k)[y(k) - D\hat{x}(kT^-)] \tag{6}$$

$$P(kT) = P(kT^-) - W(k)S(k)W(k)' \tag{7}$$

where $S(k) = DP(kT^-)D' + V(k)$, and the *Kalman gain* $W(k) = P(kT^-)D'S(k)^{-1}$.

Sampled-Data H_∞ Filtering. In practical situations, the noise statistics is usually not available a priori and the Gaussian assumption in Kalman filter may not be valid. To relax such restrictions, the mini-max H_∞ strategy is adopted in the sampled-data filtering framework to achieve robustness in motion recovery.

The performance measure of sampled-data H_∞ filter is defined directly in terms of the continuous-time system state $x(t)$ and disturbance $\tilde{v}(t)$:

$$J = \frac{\|x(t) - \hat{x}(t)\|^2}{\|\tilde{v}(t)\|^2 + \|e(k)\|^2 + (x_o - \hat{x}_o)'R_0(x_0 - \hat{x}_o)} \tag{8}$$

and thus inter-sample behavior of the system is taken into account. The weighting matrix R_0 is a confidence measure of the initial state estimate $\hat{x}_o$.

Table 1. Differences between the ground truth and the DKF/CDKF/H_∞ estimated nodal positions, (mean error ± standard deviation) $\times 10^{-2}$

Method	20dB(Gaussian)	30dB(Gaussian)	20dB(Poisson)	30dB(Poisson)
DKF	1.39±0.47	1.21±0.37	2.73±0.59	2.41±0.53
CDKF	1.27±0.43	1.09±0.34	2.56±0.54	2.27±0.51
H_∞	1.29±0.44	1.18±0.37	1.30±0.42	1.21±0.37

Table 2. Comparison of average nodal positional errors from sampled-data H_∞ filter results under various types of noise, (mean error ± standard deviation) $\times 10^{-2}$

Noise Type (20dB)	Gaussian	Poisson	Uniform	Rayleigh	Exponential
Average Positional Error	1.29±0.44	1.30±0.42	1.32±0.43	1.33±0.45	1.29±0.41

Given a prescribed noise attenuation level $\gamma > 0$, the sampled-data H_∞ filter will search for the optimal estimate $\hat{x}(t)$ such that

$$\sup J \leq \gamma^2 \tag{9}$$

where the supremum is taken over all possible disturbances and initial states. The formulation in equation (9) guarantees the bounded estimation error over all possible disturbances, regardless of the noise statistics. As a result, the filter achieves greater robustness to disturbance variations and is well suited to such real-world problems as in cardiac image analysis, where the types and levels of system disturbances and data uncertainties are not available *a priori*.

The sampled-data H_∞ filtering algorithm for the hybrid cardiac analysis system of equations (2) and (3) is given as follows [6]:

$$\dot{\hat{x}}(t) = A(t)\hat{x}(t) + B(t)w(t) \tag{10}$$

$$\hat{x}(kT) = \hat{x}(kT^-) + P(kT)D'[y(k) - D\hat{x}(kT^-)] \tag{11}$$

where $\hat{x}(kT^-) = \lim_{\varepsilon \to 0} \hat{x}(kT - \varepsilon)$, and $P(kT)$ is the stabilizing solution to the following Riccati equation with jumps:

$$\dot{P}(t) = A(t)P(t) + P(t)A(t)' + \frac{P(t)^2}{\gamma^2} + I \tag{12}$$

$$P(kT) = P(kT^-)[I + D'DP(kT^-)]^{-1} \tag{13}$$

with the initial condition $P(0^-) = R_0^{-1}$.

The filter given above is a linear system with finite jumps, which also has an intuitively appealing structure. Between the sampling instants, the state estimate evolves according to the continuous system dynamics, and the predicted state $\hat{x}(kT^-)$ is obtained by solving the differential equation (10) on the time interval $[(k-1)T, kT]$, with the previous state estimate $\hat{x}((k-1)T)$ as initial condition. Then at the observation time $t = kT$, the new measurement $y(k)$ is used to update the estimate with the filter gain being $P(kT)D'$. $P(kT^-)$ is also obtained by solving the differential equation (12) with $P((k-1)T)$ as initial condition.

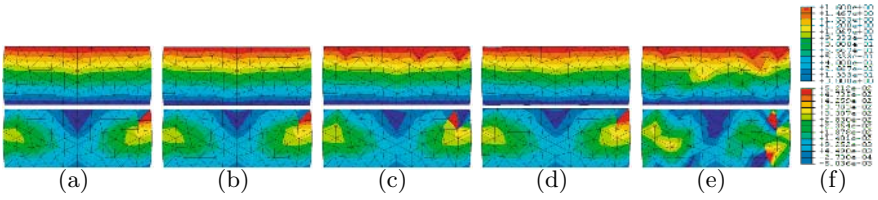


Fig. 2. Results comparison at frame #8 (the largest deformation time). Top: displacement magnitude maps. Bottom: x-strain maps. (a): ground truth. (b)(c): estimation results under Gaussian noise. (b): CDKF results, (c): DKF results. (d)(e): estimation results under Poisson noise. (d): H_∞ results, (e): CDKF results. (f): color scale.

Computational Issues. Numerical integration is usually required for solving Riccati differential equation (12), and its property of movable singularities usually leads to stability problems when Runge-Kutta routines are applied. The Möbius schemes proposed in [7] are based on viewing the Riccati equation as a flow on the Grassmannian. Since there are no singularities in the associated flow, the schemes are able to deal with numerical instability and pass accurately through the singularities. Thus we adopt Möbius schemes in our implementation, and the detailed algorithm can be found in [7]. The Lyapunov differential equation (5) is a special case of equation (12), and is solved in the same way.

2.4 Discussions

In [2, 3], the continuous-time state equation (2) in the hybrid system is first converted to discrete time and *discrete* Kalman or H_∞ filters are applied for kinematics estimation. After discretization, the state equation becomes:

$$x(k+1) = F(k+1,k)x(k) + G(k)w(k) + v(k) \tag{14}$$

where the transition matrix $F(k+1,k) = e^{AT}$, $G(k) = A^{-1}(e^{AT} - I)B$, and $v(k)$ is the discrete process noise.

There are several implicit approximations/assumptions in the above conversion. First, the assumption that the input is piecewise constant is required. Second, the transition matrix $F(k+1,k)$ in general has no explicit form, and it can be reduced to the matrix exponential form e^{AT} only when the system is time-invariant [5]. In consequence, the accuracy of the estimate from *discrete* Kalman and H_∞ filters largely depends on the extent to which the discretized state equation (14) approximates the true continuous dynamic equation (2). For large sampling interval T, the assumptions of piecewise constant input and time-invariant system become unrealistic, and thus equation (14) is not a faithful representation of the original continuous-time system dynamics.

In contrast to the system discretization approach, sampled-data filtering evolves the state estimate and the filter gain according to the original continuous system dynamics through equations (4) (5) in Kalman filter and equations (10) (12) in H_∞ filter, and thus physically more meaningful estimates could be

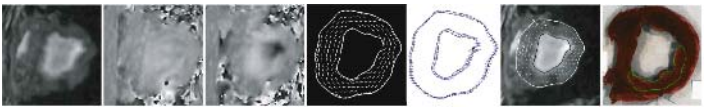


Fig. 3. Canine MRI phase contrast data (left to right): MR intensity, x-velocity, y-velocity, myofiber orientations, displacement constraints on the boundaries, phase contrast velocity field, and TTC-stained post mortem myocardium with infarcted tissue highlighted (providing the clinical gold standard for assessing image analysis results).

obtained for the continuous cardiac dynamics. And as the approximation errors in the discretization are avoided, the estimation results would be more accurate.

Sampled-data Kalman and H_∞ filters are closely related to each other. The state prediction equations (4) and (10) in the two filters are essentially the same, which is consistent with the idea that the state estimates should evolve according to the original continuous dynamics between observations, regardless of the optimization criterion. Moreover, as $\gamma \to \infty$, the Riccati equation (12) becomes the Lyapunov equation (5) despite the different weighting matrix $\tilde{Q}(t)$, and thus Kalman filter can be regarded as a limiting case of H_∞ filter when $\gamma \to \infty$. As a result, the estimation error under the worst disturbances is unbounded in Kalman filter, making it sensitive to parameter variations and noise statistics violations. In contrast, H_∞ estimation minimizes the maximum error energy over all possible disturbances, thus demonstrating greater robustness.

3 Experiments and Validation

3.1 Validation with Synthetic Data

Experiments are conducted on an object undergoing deformation in the vertical direction with the bottom being fixed. In Fig.1(a), the object is composed of two materials with different fiber orientations, and the Young's modulus $E_{fiber} = 75$, $E_{crossfiber} = 25$. Sixteen frames of the cyclic motion are acquired (Fig.1(b)), and the boundary nodes displacements are added with different types and levels of noise to generate the partial and noisy measurements. Three estimation strategies: continuous-discrete Kalman filter (CDKF), discrete Kalman filter (DKF), and sampled-data H_∞ filter (H_∞), are implemented for motion recovery. Fig.2 shows the estimated displacement magnitude and strain maps at frame #8.

The advantages of sampled-data filtering are validated through the comparison between CDKF and DKF in Fig.2(b)(c) and Table.1. CDKF results are overall closer to ground truth and smoother than DKF results, validating our analysis in section 2.4. Without system discretization, CDKF avoids the approximation errors and yields more accurate estimates under the same noise level.

The robustness of H_∞ strategy is validated through the results in Fig.2(d)(e), where the noise is Poisson distributed. For quantitative comparison, Table.1 and 2 present the point-by-point positional errors under noises of different types and levels. Overall, CDKF performs poorly if the Gaussian noise assumptions are

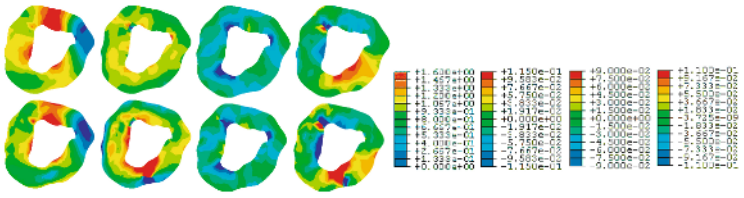


Fig. 4. Estimated displacement magnitude, radial (R), circumferential (C), and R-C shear strain maps for frame #9 (with respect to frame #1). Top: CDKF results. Bottom: H_∞ results. Right: color scales.

violated, while H_∞ filter produces accurate and very similar results for data contaminated by different types and levels of noise.

3.2 Canine Image Application

For the MR phase contrast data in Fig.3, myocardial boundaries and frame-to-frame boundary displacements are extracted using a unified active region model strategy [8], and the displacements are used as our measurements along with the mid-wall phase contrast velocities. The myofiber orientations are mapped onto the particular image slice, based on the principal warps algorithm of the landmarks. Fig.4 presents the recovered displacement and strain maps from sampled-data H_∞ filter and CDKF. The lower-right quarter of the myocardium shows much larger strain than other normal tissues, especially in the RC strain maps. These patterns are in good agreement with the highlighted TTC stained tissue in Fig.3. The computational cost is hindering the experiments for 3D cases, and we are looking into possibilities such as modal decomposition to tackle the problem.

References

1. Frangi, A., Niessen, W., Viergever, M.: Three-dimensional modeling for functional analysis of cardiac images: A review. IEEE TMI **20(1)** (2001) 2–25
2. Shi, P., Liu, H.: Stochastic finite element framework for simultaneous estimation of cardiac kinematic functions and material parameters. Medical Image Analysis **7** (2003) 445–464
3. Liu, H., Shi, P.: Cardiac motion and elasticity characterization with iterative sequential H∞ criteria. In: MICCAI. (2004) 34–42
4. Glass, L., Hunter, P., McCulloch, A.: Theory of heart. Springer (1991)
5. Bar-Sharlom, Y., Li, X., Kirubarajan, T.: Estimation with applications to tracking and navigation. Wiley (2001)
6. Sun, W., Nagpal, K.M., Khargonekar, P.: H∞ control and filtering for sampled-data systems. IEEE TAC **38** (1993) 1162–1175
7. Schiff, J., Shnider, S.: A natural approach to the numerical integration of riccati differential equations. SIAM Journal on Numerical Analysis **36** (1996) 1392–1413
8. Wong, L., Liu, H., Sinusas, A., Shi, P.: Spatio-temporal active region model for simultaneous segmentation and motion estimation of the whole heart. In: IEEE VLSM. (2003) 193–200

HMM Assessment of Quality of Movement Trajectory in Laparoscopic Surgery

Julian J.H. Leong, Marios Nicolaou, Louis Atallah, George P. Mylonas, Ara W. Darzi, and Guang-Zhong Yang

Royal Society/Wolfson Medical Image Computing Laboratory & Department of Biosurgery and Surgical Technology, Imperial College London, London, United Kingdom
{j.leong, m.nicolaou, l.atallah, george.mylonas, a.darzi, g.z.yang}@imperial.ac.uk

Abstract. Laparoscopic surgery poses many different constraints to the operating surgeon, this has resulted in a slow uptake of advanced laparoscopic procedures. Traditional approaches to the assessment of surgical performance rely on prior classification of a cohort of surgeons' technical skills for validation, which may introduce subjective bias to the outcome. In this study, Hidden Markov Models (HMMs) are used to learn surgical maneuvers from 11 subjects with mixed abilities. By using the leave-one-out method, the HMMs are trained without prior clustering subjects into different skills levels, and the output likelihood indicates the similarity of a particular subject's motion trajectories to the group. The experimental results demonstrate the strength of the method in ranking the quality of trajectories of the subjects, highlighting its value in minimizing the subjective bias in skills assessment for minimally invasive surgery.

1 Introduction

Minimally Invasive Surgery (MIS) was first introduced over 20 years ago [1-2], however, many surgeons are still restricted to performing relatively simple procedures, for example laparoscopic cholecystectomy and diagnostic arthroscopy. The uptake of advanced MIS procedures, such as laparoscopic colectomy for cancer [3], is still very slow in many countries. These procedures are mainly reserved for tertiary referral centers, performed by highly sub-specialized surgeons.

The constraints in the MIS environment have been well documented; the lack of 3D vision, limited haptic feedback and the fulcrum effect restrict the variety of surgery performed. It is, however, quite clear that some surgeons are superior in performing these tasks than others. This has motivated extensive research into the objective assessment of surgical skills. The methodology has now evolved from subjective qualitative assessment by the trainers, knowledge assessment using postgraduate examinations, to objective quantitative approaches using time or movement parameters. Quantitative methods for assessing surgical dexterity have been widely validated for a number of open and laparoscopic procedures [4]. The validation for these methods, however, relies upon prior definition of expertise, and this classification is mostly based on the assumption that experience equals technical excellence.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 752 – 759, 2006.

The purpose of this study is to examine the use of Hidden Markov Models (HMMs) based on view invariant trajectory representations for assessing complex MIS tasks. In order to increase the complexity of hand-eye coordination, view rotation tasks have been introduced in this study. This effect has been studied previously and the ability to handle mental rotation tasks has been suggested to be indicative of the innate ability in mastering laparoscopy [5-6]. We demonstrate that with the proposed method, a probabilistic framework can be formulated to allow the observation of trajectories without prior, arbitrary classification of the subjects' abilities, thus minimizing subjective bias in MIS skills assessment.

2 Modeling Instrument Motion Trajectory

2.1 Instrument Tip Tracking and Calibration

In order to obtain the trajectory of the instrument tips in Euclidean space and not interfere with the experimental task, a tracking device was attached rigidly to the handles of the laparoscopic instruments. For accurate positioning of the tracking device, a Polaris (Northern Digital Inc, Ontario) 6-DOF infrared (IR) tracker was used. The Polaris tracker is able to track a number of passive, active, wired and wireless IR tools in real time simultaneously. Data interfacing was achieved through RS-232 and the provided tracking accuracy is 0.35 mm RMS at a sampling rate of 60Hz. The offset of the instrument tips from the IR markers was calculated using the Pivot function in the NDI ToolViewer Software version 3.02.01.

The accuracy of the instrument tip tracking arrangement was further assessed by mounting the laparoscopic instrument on a Stäubli RX60L robotic arm with repeatability accuracy of +/- 0.02mm and six degrees of freedom. Ten points for calibration were set up using the robotic arm to manipulate the instrument whilst the IR markers were being tracked.

2.2 Subjects and Experimental Setup

Eleven subjects were recruited for the study (9 medical student and 2 practicing surgeons, 2 subjects were left-handed). None of the medical students had previous simulated or real laparoscopic experience. They were then randomized into two groups. All subjects were consented prior to the study, and were given a short introduction of laparoscopic surgery and the instruments used. Both groups were required to perform a laparoscopic task in a box trainer, after familiarization with the instruments and environment. A Karl Storz laparoscopic stack with an S1 camera head, Xenon Nova light source, and Hopkins II 0º endoscope was used for the experiment, along with two laparoscopic graspers with IR tracking devices rigidly attached.

For both groups, the first task was to touch two standardized points (A and B) on a simulated plastic small bowel model as illustrated in Fig. 1. Each point was attached to an in-house designed touch sensitive circuit switch to mark the beginning and end of each trajectory. Each time the circuit was completed, an alarm will indicate a successful contact. The subjects were then asked to touch alternatively the points A and B with the left instrument 10 times, and then the right instrument, and this step was repeated. The subjects were then required to perform a complex task: first, lift a

simulated omental flap using the left instrument to grasp the left corner, and touch a third point (C) hidden underneath the flap with the right. This was repeated 3 times. A total of 36 trajectories were obtained between the two points. For the second task, Group 1 were required to repeat the first task with the laparoscopic camera rotated 90 degrees counter clockwise and Group 2 with the camera rotated 90 degrees clockwise.

Fig. 1. (a) Experiment setup showing the arrangement of the IR sensors in relationship to the laparoscope tools, and the plastic small bowel model with a simulated omental flap with and without the camera view being rotated (b-c)

For the analysis of the trajectories, the experiments were divided into: (1) left instrument motion from point A to B; (2) left instrument motion from point B to A; (3) right instrument motion from point A to B; (4) right instrument motion from point B to A; (5) left instrument motion in the complex task; (6) right instrument motion to touch point C.

2.3 View Invariant Representation of 3D Trajectories

Prior to HMM analysis, the 3D trajectories were mapped to a view invariant representation based on the Centroid Distance Function (CDF) [7]. The instrument tip positions from IR tracking after offset correction were modeled as a parametric curve:

$$\begin{aligned} r[t] &= \{x[t], y[t], z[t]\} \\ t &= 0...N-1 \end{aligned} \tag{1}$$

CDF is a feature that is affine invariant, and is also widely used in image retrieval applications [7]. In essence, CDF describes a time series of the distance of each point in the trajectory from the centroid of the trajectory, and therefore after scale normalization it is rotational and translational invariant. This significantly simplifies the subsequent HMM classification by foregoing cumbersome preprocessing steps. Fig. 2 shows the CDF projection of the instrument trajectories for a surgeon and a novice, respectively.

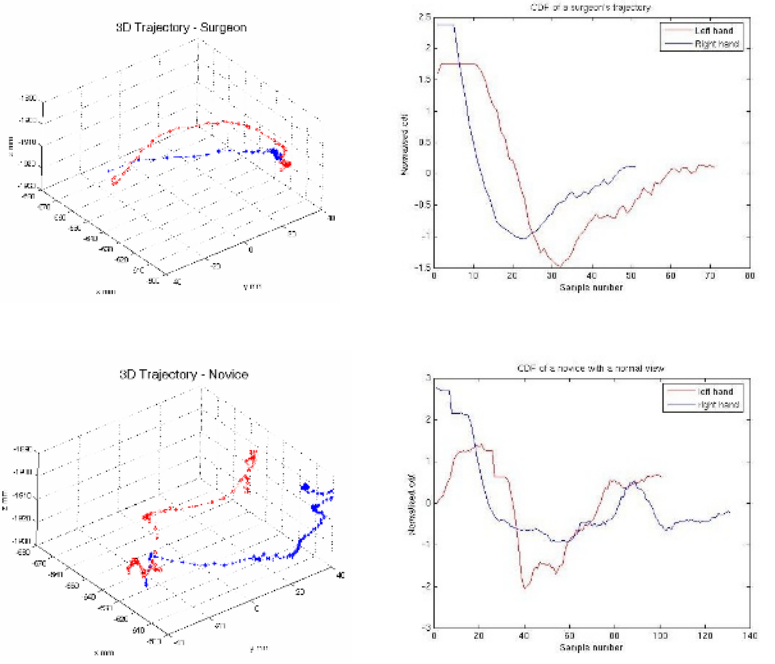


Fig. 2. Left column: 3D trajectories of surgeon and novice, blue shows the right hand and red shows the left. Right column: CDF representation of the same trajectories

2.4 Hidden Markov Modeling

HMMs are finite state stochastic machines that allow dynamic time warping for modeling time series data. HMMs have been used to classify movement trajectories, however, segmentation was necessary to avoid violating the Markovian property which assumes independence of a current state from past states given the previous one [8]. Each trajectory was regarded as one independent action, and the CDF for the trajectory was used as input signals to the HMM, hence the notion was made that each trajectory adhere to the Markovian assumption. In this study, HMM was used to learn each trajectory of a given experiment and view rotation. The leave-one-out method was used to train the HMM from all subjects excluding the test subject's data. The trained HMM was then used to calculate the log likelihood of the test subject and indicate similarities or differences to the learned model.

An HMM can be described by three model parameters representing the relationship between the hidden states (h) and the observed data (x). These parameters are:

$$(\pi_i , a_{ij} , p (x / h)) \tag{2}$$

where π_i is the initial state probability, a_{ij} the transition matrix between the hidden states and $p(x/h)$ the probability of generating an observation given the hidden state.

In this study, a fixed number of states were used. However, in order to have a more flexible model, the observation probabilities were modeled by a Gaussian Mixture Model (GMM). Thus $p(x/h)$ can be defined as:

$$p(x/h) = \sum_{m=1}^{M} c_m \frac{1}{(2\pi)^{P/2} \left|\Sigma_m\right|^{1/2}} \exp\left(-\frac{(x-\mu_m)^T \Sigma_m^{-1} (x-\mu_m)}{2} \right) \quad (3)$$

where c_m is the mixing parameter, μ_m and Σ_m are the mean and covariance matrix of the component m of the GMM.

The K means algorithm was used to initialize the parameters of the observation GMM (mainly the means μ_m). Two versions of the K means algorithm were implemented to compare performance, using Euclidean distance and Derivative Dynamic Time Warping [9] to calculate the similarity between two trajectories. The second option can take signal 'warping' into consideration when finding cluster centers, and the length of the 'warping' was then used to normalize the distance between the two trajectories. As the resulting cluster centers were identical with both methods, Euclidean distance was used as it was more computational efficient. The Expectation Maximization (EM) algorithm was used to calculate the maximum likelihood of the parameters of the HMM, namely the means and covariances of the components of the GMM and the parameters π_i and a_{ij}.

2.5 Categorized Observational Score

The video of all the tasks were scored by two independent observers who were blinded to the identity of the subjects, the scoring system used was a modified version of the Objective Structured Assessment of Technical Skills (OSATS) global rating scale. This is a widely validated score developed by Martin *et al*, using 8 categories each with a Likert scale of 1-5 anchored by descriptors [10]. The modification was necessary as 3 of the categories did not apply to this particular experiment (suture handling, knowledge of procedure and quality of final product).

3 Results

To demonstrate how the effect of view rotation increases the complexity of the tasks, Fig. 3(a) illustrates the average time increased for a subject to complete a trajectory when the view is rotated. Fig. 3(b) demonstrates the OSATS global rating scale (with good inter-observer agreement) decreases with view rotation, however this does not completely correlate with the time measurement. For example, subject 2 in the rotated task had the third lowest score in OSATS, but had the fifth longest time. OSATS is an assessment of the quality of technique, whereas time is a crude metric for performance. There is a significant correlation between mean (unfiltered) path length and time taken ($r = 0.916$ $p<0.001$). Interestingly, one subject improved slightly in the rotated task (subject 7).

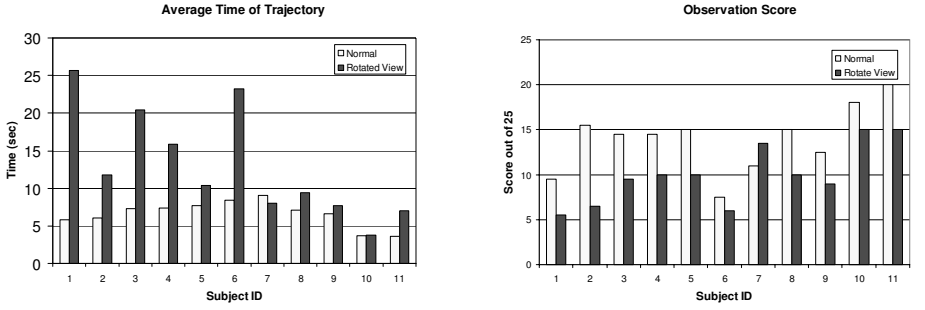


Fig. 3. (a) The average time for the trajectories of each subject. (b) The modified OSATS score for all the subjects involved in the experiments.

CDF representations of motion trajectories are illustrated in Fig. 4(a). This figure shows that the surgeon's trajectory generally lies closer to the mean of all the CDF than the novice's trajectory in this particular experiment. However, the average CDF trajectory has a distinctively different shape to either of the sample trajectories.

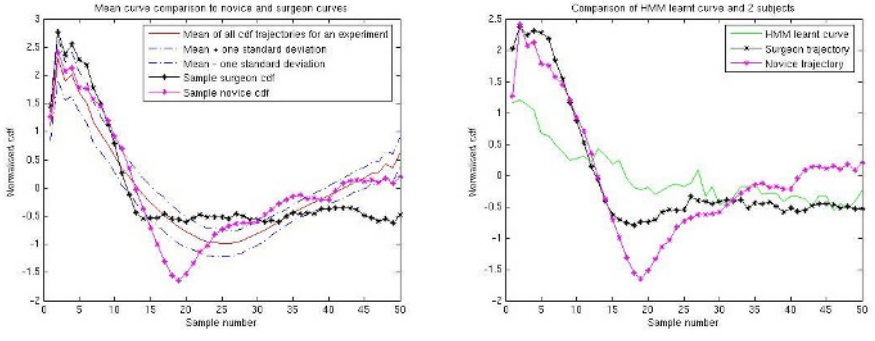


Fig. 4. (a) The mean of CDF trajectory for one of the experiments with standard deviation. An example of a practicing surgeon's trajectory is shown in black and a novice in pink. (b) HMM learned curve shown in green.

In Fig. 4(b), the HMM learnt curve is plotted instead, this is the average of ten observations after training the HMM. The set of trajectories showed a large variability with effects of time warping. By calculating the likelihood to the HMM, a measure of similarity of a certain trajectory to the whole group can be obtained.

The parameters of the HMM including the number of Gaussians in the observation GMM, as well as the number of the hidden states of the HMM were selected experimentally. Parameters that lead to the least variation in the values of the test data likelihoods were selected as the parameters that can provide good data representation for this dataset. The number of hidden nodes was selected as 4 with a mixture of 2 Gaussians for the GMM.

The log likelihoods of the subjects' trajectories belonging to the training data sets, using the leave-one-out method are illustrated in Figs. 5(a) and (b). Lower values

indicate the subjects' trajectories are more likely to match the models learnt by the HMMs. The log likelihood values are negative. The effect of view rotation accentuates the difference between the subjects as shown in Fig. 5(b).

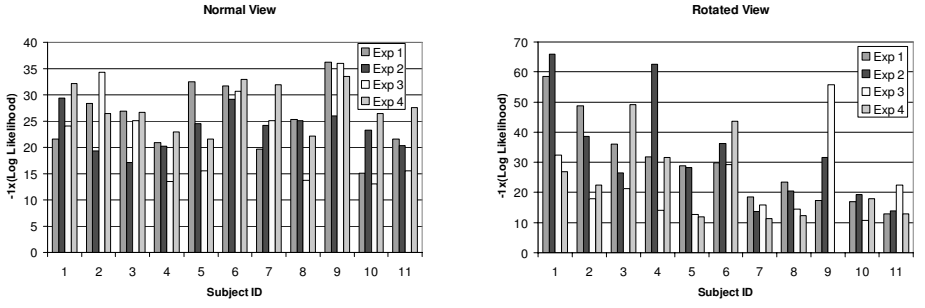


Fig. 5. (a) The negative log likelihood of the subject in each experiment to belong to the group in the normal view orientation and (b) rotated

Finally, the log likelihood values were ranked in each experiment independently, rank number 1 represents the most similar and vice versa. The mean rank of experiment 1-4 for each subject was calculated in both normal and rotated view in Fig 6(a). In order to compare this with the most validated surgical rating scale, this rank is plotted against the ranking in the OSATS score in Fig. 6(b). There was a very significant correlation between them ($r=0.93$ $p<0.001$).

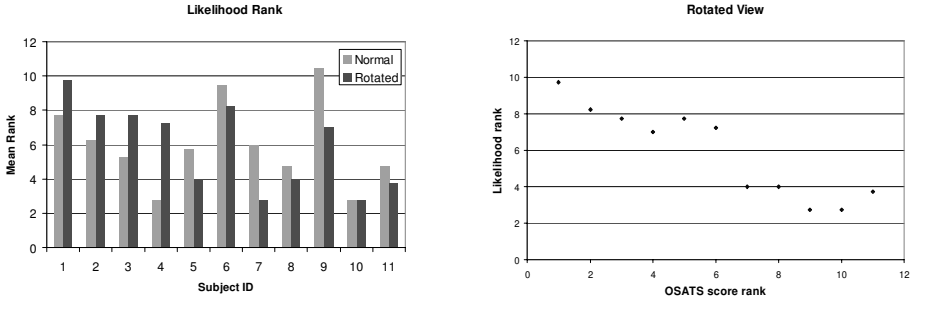


Fig. 6. (a) The mean rank of subjects' likelihood of belonging to the test group, (b) scatter plot of the rank of likelihood generated by the trained HMM against OSATS score ranking in the rotated tasks

4 Discussion and Conclusions

This study has shown that HMM can be used to learn models of surgical motion trajectories even in a group of subjects with mixed abilities without prior classification of technical skills levels. The HMMs calculated the likelihood of the practicing surgeons as the most representative trajectories (subjects 10 and 11 in figure 6), this also

correlated very well with the observation scores, which is a labor intensive method to assess the quality of surgical technique. One of the important considerations of applying the proposed HMM scheme is the feature representation of the trajectories. In general, it should be invariant to affine transformation, as this can cope with trajectories with different starting points, rotations and approaching directions. In this study, we have used CDF as a means of invariant feature representation. Other approaches based on extrema in acceleration measured by high frequency wavelet coefficients are also applicable [11].

References

1. Litynski G.S.: Profiles in laparoscopy: Mouset, Dubois and Perissat: the laparoscopic breakthrough in Europe (1987-1988). JSLS **3(2)** (1999) 163-167
2. Semm K.: Endoscopic appendectomy. Endoscopy **15(2)** (1983) 59–64
3. Harinath G., Shah P.R., Haray P.N., Foster M.E.: Laparoscopic colorectal surgery in Great Britain and Ireland--where are we now? Colorectal Dis **7(1)** (2005) 86-89
4. Moorthy K., Munz Y., Sarker S.K., Darzi A.: Objective assessment of technical skills in surgery. BMJ **327(7422)** (2003) 1032-1037
5. Jordan J.A., Gallagher A.G., McGuigan J., McClure N.: Randomly alternating image presentation during laparoscopic training leads to faster automation to the "fulcrum effect". Endoscopy **15(10)** (2000) 317-321
6. Nicolaou M., James A., Darzi A., Yang G.Z.: Extraction of laparoscopic strategies of novices using eye tracking. In MIPS XI conference (Windermere, UK, 2005)
7. Bashir F., Khokhar A., Schonfeld D.: View-Invariant Trajectory-Based Activity Classification and Recognition. ACM Multimedia Systems Journal (MMSJ) (2006)
8. Rabiner L.R.: A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of IEEE **77(2)** (1989) 267—296
9. Keogh E.J. and Pazzani M.J.: Derivative Dynamic Time Warping. In First SIAM International conference on Data Mining (Chicago, IL, 2001)
10. Martin J.A., Regehr G., Reznick R., MacRae H., Murnaghan J., Hutchison C., Brown M.: Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg. **84(2)** (1997) 273-278
11. Chen W., Chang S.F.: Motion Trajectory matching of Video Objects. Proceedings of SPIE **3972** (2000) 544-553

A Novel MRI Texture Analysis of Demyelination and Inflammation in Relapsing-Remitting Experimental Allergic Encephalomyelitis

Yunyan Zhang[1,2], Jennifer Wells[2], Richard Buist[3], James Peeling[3], V. Wee Yong[2], and J. Ross Mitchell[1,2]

[1] Department of Radiology, University of Calgary, Calgary, Alberta, CA T2N 1N4
{yunyzhan, rmitch, vyong}@ucalgary.ca
[2] Department of Clinical Neurosciences, University of Calgary, Calgary, AB, CA T2N 1N4
JELA@lundbeck.com
[3] Department of Radiology, University of Manitoba, Winnipeg, MB, CA R3T 2N2
{rbuist, jpeeling}@ms.umanitoba.ca

Abstract. We have developed a novel multiscale localized image texture analysis technique, based upon the polar Stockwell Transform (PST). In this paper we characterized image texture in vivo using the PST in histologically verified lesion areas in T2-weighted MRI of an animal model of multiple sclerosis. Both high and low frequency signals, representing inflammation and demyelination, were significantly increased in pathological regions compared to normal control tissue. This suggests that this new local spatial-frequency measure of image texture may provide a sensitive and precise indication of disease activity.

1 Introduction

Multiple sclerosis (MS) is a chronic inflammatory demyelinating disorder of the central nervous system (CNS). This disease affects over one million people worldwide, and is the leading cause of disability in young adults.[1] Magnetic resonance imaging (MRI) is a well-established imaging modality for diagnosis and monitoring MS.

The characteristic focal "plaques" in the CNS of MS patients are readily demonstrated as hyperintense areas on T2-weighted MRI. However, subtle diffuse pathological changes in normal appearing white matter (NAWM)[2] are difficult to detect using conventional MRI techniques. Nevertheless, these abnormalities are believed to play a pivotal role in the neurological impairments in patients with MS.

The infiltration of inflammatory cells, including macrophages, predominates the cascade tissue damage during MS pathogenesis. Infiltrated macrophages can be labeled by uptaking systemically administered ultra small particles of iron oxide (USPIO) via phagocytosis.[3] USPIO loaded macrophages at sites of inflammation induce significant shortening of MRI T2 relaxation time, which leads to decreased signal intensity on T2-weighted MRI.[4] Therefore, conventional MR imaging can be complemented by the tracking of macrophages *in vivo* in the study of MS.

We hypothesize that as tissue becomes abnormal, its MRI 'texture' will change. Image texture refers to a local characteristic pattern of image intensity that identifies a tissue. Texture analysis generates computer measures that quantitatively describe

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 760 – 767, 2006.

texture content in an image. A majority of published accounts have used statistical techniques, the co-occurrence matrix method[5] for instance, to analyze image texture in MS[6-8], and other diseases.[9] In particular, statistical texture analysis has been applied with promise to: distinguish active and inactive lesions; differentiate normal and pathological spinal cord; and, characterize therapeutic response in T2-weighted MRI in MS. However, current statistical methods are limited in detecting subtle abnormalities in the NAWM in MS.[7]

The polar Stockwell Transform (PST) is a new analysis technique that represents texture by its local spatial frequency. The PST combines features of Fourier and wavelet transforms to provide both spatial and spectral information throughout an image.[10,11] That is, it uses a frequency-dependent localizing Gaussian window to isolate the Fourier domain for sub-region analysis. Thus, the PST is capable of detecting subtle characteristics of image pattern and intensity. Indeed, previous MRI PST studies[12] in MS showed that low spatial frequency content increased in an area of NAWM up to 30 days prior to lesion formation, and decreased when active lesions became inactive. The in-vivo human studies, however, were unable to provide histological correlation of the MRI texture changes.

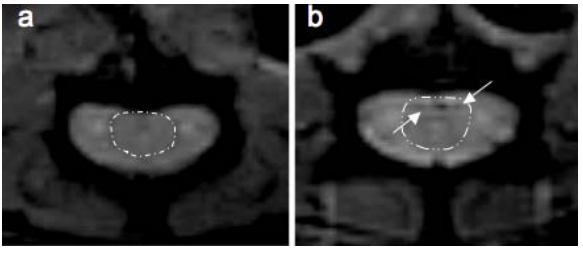


Fig. 1. Example 7 T T2-weighted MRI from a control (left) and an EAE (right) mouse. EAE cord (dashed oval) had higher signal intensity, larger volume than control cord. Two small low signal intensity lesions (arrows) were identified at the ventral side of lumber EAE cord.

The goal of the present study was to elucidate the histological substrate of image texture analyzed in the USPIO enhanced MRI from a murine model of MS, experimental allergic encephalomyelitis (EAE). The PST was applied to histologically defined lesion areas on 7 T T2-weighted MRI from the spinal cord of mice with EAE. This provided a unique opportunity to investigate local MRI spatial frequency content and compare that to tissue histopathology.

2 Materials and Methods

2.1 Induction of EAE

Eight to ten week old strain-129 mice were studied. EAE was induced in four mice by injecting subcutaneously 50 μg of the immunogenic myelin oligodendrocyte glycoprotein (MOG) peptide (amino acids 35-55, synthesized by the Peptide Facility of the University of Calgary) emulsified in 100 μl of complete Freund adjuvant (CFA)

(Difco Laboratories, Detroit, Michigan). Pertussis toxin (0.3 μg/200 μl, List Biological Laboratories, Campbell, Calif., USA) was injected intraperitoneally on the same day, and again two days later. EAE induced this way in stain-129 mice is characterized by relapsing-remitting episodes. Mice were weighed and evaluated daily. The degree of disability caused by EAE was assessed using neurological scores graded from 0 (normal) to 15 (death from EAE). Two age and strain matched non-EAE mice served as controls. All protocols for animal research conformed to the guidelines of the Animal Research Council of the Universities of Calgary and Manitoba as well as the Canadian Council on Animal Care.

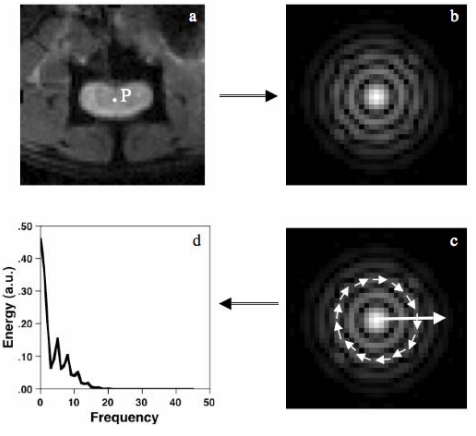


Fig. 2. The 1D spectral curve (d) was obtained by collapsing the 2D PST (b) calculated for pixel P (a)

2.2 MRI Experiments

After induction of EAE, *in vivo* MR scans were performed at different time points, recorded as days post immunization (dpi). Two mice were imaged during remission of EAE (38 dpi and 21 dpi) and two were imaged at disease peak during the first relapse (35 dpi and 49 dpi). Inhalational isoflurane anesthesia (1.5% in oxygen:N_2O 30:70) was delivered via a nose cone during the imaging procedure. Each animal was positioned in a supine position within a 1.9 cm diameter birdcage coil. Respiration was monitored and the body temperature was maintained at 37^0C by a thermocouple-based system during the MRI experiments.

The MRI examinations were performed on a 7 T Bruker horizontal bore NMR spectrometer. Ferumoxytol (AMI7228, Advanced Magnetics, Cambridge, MA) USPIO (30mg Fe/kg) was administered 24 hours before MRI scans to allow time for the USPIO-loaded macrophages to move into inflammation sites. T2-weighted MRI was acquired using an 8-echo pulse sequence (TR/TE = 2540ms/26.79ms, matrix size = 256 x 256, FOV = 2.5 x 2.5 cm^2, slice thickness = 0.75 mm) (Fig.1). This provided 12 interleaved slices covering the spinal cord from the lumbar (L2) to sacral segment (9 mm in length). For comparison, two control mice (without EAE, with USPIO) were also imaged using the same protocol.

2.3 Histopathology Examinations

Immediately following the imaging experiment, the anesthetized animal was euthanized. The imaged spinal cord was carefully removed and immersed in 10% buffered formalin. Then the 9 mm long spinal cord segment was dissected into 6 1.5 mm blocks. Each of these blocks was then embedded in paraffin. Anatomical landmarks were strictly followed during this process to ensure the location of cord dissection matched with MRI slice position. Two continuous axial 6 μm thick sections were cut every 0.75 mm and placed on separate glass slides. Two series of 12 sections each were prepared per animal, for a total of 144 slide preparations ([4 EAE + 2 control] mice x 2 series x 12 sections). The first series of sections was stained with hematoxylin, eosin, and luxol fast blue for evidence of inflammation and demyelination. The second series was examined for the presence of macrophages and USPIO co-localized by Iab-1 immunostain and Prussian blue.

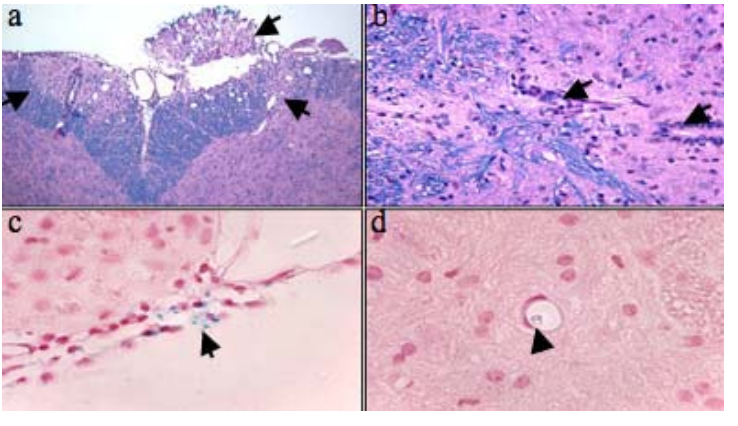


Fig. 3. Histology of the MRI lesion ROI at Fig. 1. Shown are demyelination (arrows, a), inflammation (arrows, b), and Prussian blue positive macrophages in the glia limitans (arrow, c) and in the blood vessel (arrow, d).

2.4 PST Texture Analysis

The polar Stockwell Transform of an image $I(x,y)$ is defined as follows:[10]

$$S(x,y,k_x,k_y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} I(x',y')w(x'-x,y'-y,f,\theta)e^{-i2\pi(k_x x'+k_y y')}dx'dy' \tag{1}$$

Where, $w(x,y,f,\theta)$ is the localizing Gaussian window

$$w(x,y,f,\theta)=\frac{k^2\lambda(k)}{2\pi}e^{-\frac{k^2}{2}(x\cos\theta+y\sin\theta)^2}e^{-\frac{(k\lambda(k))^2}{2}(-x\sin\theta+y\cos\theta)^2} \tag{2}$$

Here, $k = \sqrt{k_x{}^2 + k_y{}^2}$ is the centre frequency, and $\theta = a\tan(\frac{k_y}{k_x})$ is the frequency orientation. The Gaussian window is inversely proportional to the center frequency, k, and is rotated to the direction θ. The PST is rotationally invariant, and therefore eliminates the effect on texture measurements of patient or image rotation.

The obtained MR images were non-uniformity corrected prior to any image processing using a polynomial algorithm developed in house. This process removed the possible intensity bias caused by the high field MR system and greatly improved the MRI contrast. Lesion-guided analysis was employed to determine the texture characteristics of pathological tissues. That is, texture was only analyzed from the low signal intensity areas on T2-weighted MRI when pathological changes at the same location were present on the spinal cord sections at the corresponding MRI slice position.

The PST spectrum of each ROI (32x32 pixels) was computed to first generate a local 2D Fourier spectrum for each pixel. Each local 2D spectrum was then reduced to a local 1D spectrum by integrating it along rings of constant width (0.32 cm^{-1}) in the Fourier domain (Fig. 2). The local 1D spectra from the central 2x2 pixels in each ROI were averaged for analysis. An estimate of the low frequency energy within each ROI was obtained by summing the area under the local 1D spectrum below frequency = 3.2 cm^{-1}. The same ROIs in the corresponding MRI slices from normal mice were selected, and the 1D PST spectrum was calculated to obtain control values.

2.5 Statistical Analysis

The paired Student *t*-test and statistical software program (SPSS 11 for Macintosh) were used to compare MRI texture from EAE mice and from normal control mice. A *P* value of less than 0.05 was considered significant.

3 Results

3.1 MRI-Matched Neuropathology of Mice with EAE

The mice imaged during remission started to show behavioral evidence of neuropathology at 12 dpi and 14 dpi. Their disability scores on the date of MRI were 1 and 2 respectively. The EAE signs of the other two mice started relatively late (23 dpi and 34 dpi), but the disability progressed fast such that they had disability scores of 9 and 7.5 when MRI was performed.

Eight ROIs in T2-weighted MR images, acquired during symptom remission, were matched with histopathology. Within these ROIs, there was a mixture of inflammation and demyelination, with the former dominating (Fig. 3, a and b). Prussian blue positive cells (monocytes and macrophages) were detected in the cord parenchyma, blood vessels and glia limitans (Fig. 3, c and d). The EAE mice imaged during disease peak had cellular infiltration throughout the spinal cord cross-sections. Consequently, focal lesions were not visible and could not be matched to the MRI ROIs.

3.2 Lesion Texture Changes Compared to Normal Control Tissue

The spectral distributions were more variable in lesion ROIs than in normal tissue ROIs. However, all spectra had similar distributions showing greater energy at the lower frequency range ($0<f \leq 3.2$ cm^{-1}) than at the higher frequencies ($3.2<f \leq 6.72$ cm^{1}) (Fig. 4). The mean spectral distribution from lesion regions also had a similar shape but with increased energy compared to that from normal spinal cord areas.

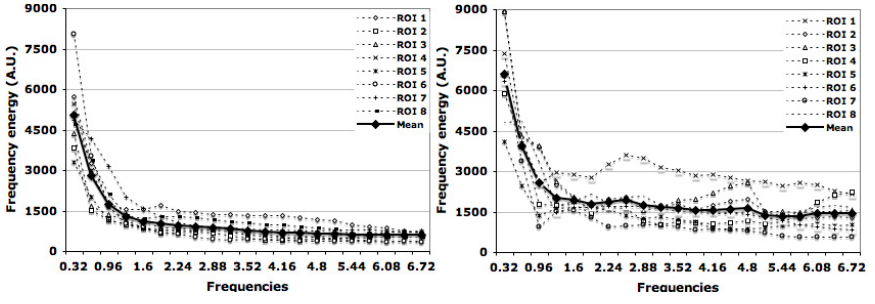


Fig. 4. Individual spectral distribution pattern and the mean spectrum of all ROIs from control (left) and EAE (right) mice quantified by the PST technique on T2-weighted MRI

The mean total energy at all spatial-frequencies was significantly higher ($p<0.01$) in lesion ROIs than in the corresponding normal tissue ROIs. The ratio of high to low frequency power was 0.63 in lesions (low, high frequency was 61.5% and 38.5%) and 0.43 in control spinal cord (low, high frequency was 69.7% and 30.3%). Further, both the low and high frequency energy was significantly increased ($p<0.01$) in diseased areas compared to normal control data (Fig. 5).

4 Discussions

This study tested the hypothesis that MRI texture, quantified by the novel local multiscale texture analysis, changes following tissue pathology. In the present study, the histological pathology was considered as the gold standard for locating the particular MRI abnormalities for texture analysis.

This study revealed that histopathological changes in EAE lesions corresponded to a significant increase in the local spectral energy determined by the PST, compared to controls. To our knowledge, this is the first study correlated tissue histopathology with MRI texture directly in vivo. A previous study,[13] using statistical texture analysis, detected remyelination in serial 4.7 T T2-weighted MR images during 2-month disease course, in a demyelinating-remyelinating cuprizone mouse model. However, no histological analysis was performed in these experiments. Instead, the authors relied upon results of the cuprizone model in the literature to infer myelin integrity for comparison. Consequently, direct correlation between image texture and tissue pathology was not possible.

Generally, low frequency components correspond to large spatial scale and coarse texture. High frequency components correspond to small spatial scale and fine

texture. Inflammatory cells, damaged myelin and fragments, and possibly axonal transection (data not shown), increased the hybridism of tissue structure. This may cause an unequal signal intensity change on T2-weighted MRI and a change in image texture. The elevated energy of the low frequency components suggests a coarse texture change in the affected tissue. On the other hand, inflammation usually accompanies edema, which first enlarges the EAE cord volume.[14] Secondly, the increased extracellular water space may smooth away the uniqueness of tissue structure. Hence the tissue becomes finer that resulted in, at least partially, increased high frequency content.

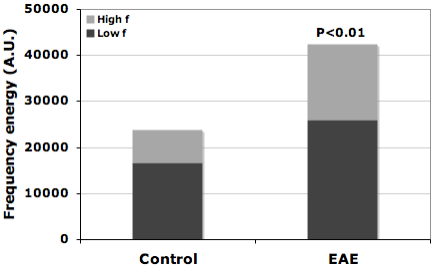


Fig. 5. Mean total spectral energy in EAE was significantly greater than that in normal tissue. The low frequency energy in pathological regions was also significantly greater than in control regions, so does the high frequency energy

Inflammatory cell infiltration is the earliest event in MS lesion initiation.[1,3] Each of the lesion ROIs examined in this study contained variable degrees of inflammation including cellular infiltration. These pathological changes, however, do not always cause visible changes in T2-weighted MRI.[15] In contrast, our new texture measure was significantly elevated in all ROIs where pathological changes were noted, even those with subtle and small histological changes. These findings are similar to those of a previous PST study in T2-weighted MRI in MS lesions, where the local frequency energy increased significantly in NAWM one month before lesion visualization.[12] This suggests that the PST is sensitive to subtle texture changes even in the early stage of lesion formation.

Cellular infiltration inevitably causes subsequent demyelination during disease process. As a consequence, almost all the histology sections contained a mixture of inflammation and demyelination. This indicates that inflammation and demyelination each contributed to the spectral energy increase in lesion ROIs. Therefore, the PST analysis is nonspecific. However, the existence of either pathology in the tissue may induce texture changes in MRI, which can be readily captured by the PST.

A few MR imaging techniques have emerged in recent years to detect subtle histopathological changes in pre-lesion areas in normal appearing brain tissue but none of them is conclusive. The PST has been shown of potential to be a sensitive technique for MRI texture analysis in MS. Therefore, determining the histopathological correlates underlying PST texture is critical for its precise interpretation and accurate application, particularly in longitudinal MS studies.

In summary, this study showed that the pathological changes in an animal model of MS caused significant increase in both low and high frequency energy characterized by the PST. In this paper, a direct correlation of tissue histology with a PST analysis

of MRI texture was used. It provided insight into the histopathological substrates causing local MRI spatial-frequency changes in MS. Hence it is of importance in directing future texture analysis in MS using the PST. Because of its sensitivity to image intensity and pattern changes, the PST may become a powerful tool to characterize subtle abnormalities in NAWM prior to lesion formation in MS. Furthermore, it may also aid in predicting early therapeutic responses in MS clinical trials.

Acknowledgements

We thank Mr. Brian O Brien for the development of MRI non-uniformity correction algorithm, and Dr. Hongmei Zhu for the initial implementation of the PST program.

References

1. I. Pirko, J. Gamez, A. J. Johnson, S. I. Macura, and M. Rodriguez. Dynamics of MRI lesion development in an animal model of viral-induced acute progressive CNS demyelination. *NeuroImage* 2004, 21: 576-582.
2. D. H. Miller, A. J. Thompson, M. Filippi. Magnetic resonance studies of abnormalities in the normal appearing white matter and grey matter in multiple sclerosis. *J Neurol.* 2003, 250: 1407-1419.
3. S. Floris, E. L. A. Blezer, G. Schreibelt, E. Dopp, et al. Blood-brain barrier permeability and monocyte infiltration in experimental allergic encephalomyelitis. A quantitative study. *Brain*, 2004, 127: 1-12.
4. V. Dousset. In vivo Macrophage activity imaging in the central nervous system detected by magnetic resonance. *MRM* 1999, 41: 329-333.
5. R. M. Haralick, K. Shanmugam, and J. Dinstein. Texture features for image classification. *IEEE Trans Syst Man Cybern.* 1973, 6: 610-621.
6. O. Yu, Y. Mauss, G. Zollner, I. J. Namer, J. Chambron. Distinct patterns of active and non-active plaques using texture analysis on brain NMR images in multiple sclerosis patients: preliminary results. *MRI* 1999, 17: 1261-1267.
7. J. M. Mathias, P. S. Tofts. N. A. Losseff. Texture analysis of spinal cord pathology in multiple sclerosis. *MRM* 1999, 42: 929-935.
8. Y. Zhang, H. Zhu, R. Ferrari, W. Wei, et al. Texture analysis of MR images of minocycline treated MS patients. *MICCAI proceeding* 2003, 2878: 786-793.
9. R. A. Lerski, K. Straughan, L. R. Schad, D. Boyce, et al. MR image texture analysis – an approach to tissue characterization. *MRI* 1993, 11: 873-887.
10. R. G. Stockwell, L. Mansinha, and R. P. Lowe. Localization of the complex spectrum: The S transform. *IEEE Trans. Signal Process*. 1996, 44: 998-100.
11. H. Zhu, B. G. Goodyear, M. L. Lauzon, R. A. Brown, et al. A new local multiscale Fourier analysis for medical imaging. *Med. Phys.* 2003, 30: 1134-1141.
12. Y. Zhang, H. Zhu, L. M. Metz, J. R. Mitchell. A new MRI texture measure to quantify MS lesion progression. *12th ISMRM proceedings*, 2004.
13. O. Yu, J. Steibel, Y. Mauss, B. Guignard, B. Eclancher, et al. Remyelination assessment by MRI texture analysis in a cuprizone mouse model. *MRI* 2004, 22: 1139-1144.
14. Y. Zhang, J. Wells, R. Buist, X. Sun, et al., "In vivo MRI quantification of mouse spinal cord with Relapsing Remitting EAE: Disability and imaging correlates," *13th ISMRM proceedings*, 2005.
15. L. L. Cook, P. J. Foster, J. R. Mitchell, and S. J. Karlik. In vivo 4.0-T magnetic resonance investigation of spinal cord inflammation, demyelination, and axonal damage in chronic-progressive experimental allergic encephalomyelitis. *J MRI* 2004, 20: 563-571.

Comparison of Different Targeting Methods for Subthalamic Nucleus Deep Brain Stimulation

Ting Guo[1,3], Kirk W. Finnis[4], Sean C.L. Deoni[5], Andrew G. Parrent[2], and Terry M. Peters[1,3]

[1] Robarts Research Institute
[2] The London Health Sciences Centre, London, ON, Canada N6A 5K8
[3] Biomedical Engineering, University of Western Ontario, London, ON, Canada N6A 5B9
[4] Atamai Inc, London, Ontario, Canada N6B 2R4
[5] Centre for Neuroimaging Sciences, Institute of Psychiatry, King's College, London, UK
{tguo, tpeters}@imaging.robarts.ca

Abstract. The subthalamic nucleus (STN) has been adopted as a commonly used surgical target in deep brain stimulation (DBS) procedures for the treatment of Parkinson's disease. Many techniques have been developed to facilitate STN DBS targeting, and consequently to improve the surgical outcome. In this work, we conducted a retrospective study on 10 patients who were treated with bilateral STN DBS to assess the target localization accuracy and precision of six methods in STN DBS surgery. A visualization and navigation system integrated with normalized functional and anatomical information was employed to perform the targeting procedures. Actual surgical target location determined by an experienced neurosurgeon with pre-operative image-guided surgical target/trajectory planning and intra-operative electrophysiological exploration and confirmation was considered as the "gold standard" in this evaluation and was compared with those localized using each of the six targeting methods. The mean distance between the actual surgical targets and those planned was 3.0 ± 1.3mm, 3.2± 1.1mm, 2.9 ± 1.1mm, 2.7 ± 1.2mm, 2.5 ± 1.0mm, and 1.7 ± 0.8mm for targeting approaches based on T_2-weighted magnetic resonance image (MRI), brain atlas, T_1 and T_2 maps, electrophysiological database, collection of final surgical targets of previous patients, and the combination of these functional and anatomical data respectively. The results demonstrated that the use of functional data along with anatomical data provides reliable and accurate target position for STN DBS.

1 Introduction

The subthalamic nucleus (STN), which is an obliquely oriented biconvex lens-shaped mass of grey matter, located dorsal to the internal capsule rostral to the substantia nigra within the caudal diencephalon in the ventral part of the subthalamus, plays an important role in the basal ganglia-thalamocortical motor circuit. Considered as the main "driving" factor in basal ganglia, the STN utilizes the excitatory neurotransmitter glutamate and innervates both basal ganglia output nuclei, the internal segment of the globus pallidus (GPi) and the pars reticulata of the substantia

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 768 – 775, 2006.

nigra (SNr). The generation of the abnormal patterns of activity observed in Parkinson's disease and essential tremor is thought to be caused by the hyperactivity of the STN. Therefore, interruption of STN activity by creating a small lesion or using high frequency deep brain stimulation (DBS) alleviates motor abnormalities such as tremor, rigidity, bradykinesia, postural instability, and akinesia [1].

Since the STN is surrounded by delicate deep-brain structures and its average size is very small, i.e. 5.9mm, 3.7mm, and 5mm in anteroposterior, mediolateral, and dorsoventral dimensions respectively [2], it is critical to localize the surgical target accurately with minimal invasiveness to achieve the optimal surgical outcome. In clinical practice, prior to the final lesion creation or stimulator placement, the surgical targeting is currently primarily determined by pre-operative image-guided surgical target/trajectory planning and intra-operative electrophysiological confirmation. The first phase is carried out based on pre-operative magnetic resonance (MR) or computed tomography (CT) images, whereas the second step involves invasive intra-operative electrophysiological measurements to refine the optimal surgical targets.

To enhance the targeting accuracy and reduce the invasiveness, clinicians have employed a variety of techniques including printed and digitized anatomical brain atlases [3,4], T_2-weighted MR images [5], high resolution T_1 and T_2 maps [6], electrophysiological databases [7], registered surgical targets of previous patients [8], as well as integration of multiple functional and anatomical references [9]. Anatomical atlases contain information about subsections of deep-brain nuclei, which are usually not distinguishable on regular MR images. However the atlases derived from single cadaver brain specimen or multiple subject brain hemispheres may not be able to provide sufficient statistical representation of the population and be difficult to generalize. Special MR T_1 and T_2 maps with excellent resolution and signal to noise ratio carry the disadvantage of lack of anatomical variation. Patient specific T_2-weighted MR images may allow better visualization of the STN, nevertheless the inconsistency of the target positions defined with T_2-weighted MR images and with electrophysiological explorations should be used with caution [10,11].

Stereotactic targeting of the STN using registered surgical targets and electrophysiological databases, which contains data from multiple patients normalized into a standard brain template, is more promising, because they provide valuable probabilistic maps of population-based actual target information and deep-brain electrophysiological activity. Furthermore, the integration of multiple functional and anatomical references can establish the relationship between functional organization and anatomic structures for estimating the surgical targets. In this work, we conducted a retrospective study to evaluate and compare the target localization accuracy and precision of six targeting methods in STN DBS surgery.

2 Materials and Methods

2.1 Patients and MRI Data

Patient selection: MR images of 10 patients, who were treated with bilateral STN DBS procedures to alleviate their Parkinson's disease symptoms at London Health Sciences Centre (LHSC), London, Ontario, Canada, were selected from the database

containing data of 26 (15 male and 11 female) patients (21-79 years, mean 59.7 years), who had undergone a total of 49 STN DBS procedures at LHSC.

MRI Data: We acquired pre-operative T_1-weighted MR images using a 3D SPGR sequence, and T_2-weighted MR images using a 2D fast spin-echo sequence on a 1.5T GE Signa scanner with the imaging parameters listed in Table 1.

Table 1. Imaging parameters of T_1-weighted and T_2-weighted MR images

Imaging Modality	TR (ms)	TE (ms)	Flip Angle	NEX	Voxel Size	Dimension
T_1-weighted	8.9	1.9	20°	2	$1.17\times1.17\times1mm^3$	256×256×248
T_2-weighted	2800	110	90°	4	$1.0\times1.0\times1.5mm^3$	256×256×21 (coronal)
						256×256×20 (axial)

2.2 "Gold Standard" and Surgical Procedure

Gold standard: Actual surgical target locations determined by an experienced neurosurgeon using standard surgical procedures were considered as the "gold standard" in this evaluation and were compared with those localized using each of the six targeting methods.

Surgical procedure: The Leksell G stereotactic frame (Elekta Instruments AB, Stockholm, Sweden) with attached MR compatible localizer was applied to each patient immediately before imaging. After image acquisition, the neurosurgeon would identify the internal landmarks, anterior commissure (AC) and posterior commissure (PC), relative to the frame system on the MR images, reconstruct the images in AC-PC-based coordinates, align the nearest section of an anatomical brain atlas to the images, and select the initial the surgical target. Before the final placement of the deep brain stimulator, five electrodes were applied to obtain the micro-recording and macro-stimulation measurements on a track extending 10 mm above and below the initially planned target and to eventually localize the optimal surgical position.

2.3 Visualization and Navigation System

A comprehensive visualization and navigation system integrated with normalized functional and anatomical information [9] within the standard brain space was employed to perform the surgical targeting procedures in our study. This system incorporates a digitized and segmented anatomical brain atlas [3], a 3D electrophysiological database, high resolution T_1 and T_2 maps [6], and collections of real surgical targets. It is also capable of interactively displaying the linked patient image and standard brain template, reformatting them along arbitrary axes, and registering and fusing images from different modalities, digitized brain atlases, and functional data with the pre-operative images. The system readily accommodates inter-subject anatomical variability through the use of a fast three-dimensional non-rigid image registration approach, AtamaiWarp [12]. Virtual surgical probes may also be manipulated within this environment to allow data-sampling points to be readily

visualized with respect to pre-defined targets and previously electrophysiological data imported from the database.

2.4 Targeting Techniques

At our institution, the dorsolateral portion of STN is regarded as the most effective stimulation site in STN DBS. Therefore, all the target positions estimated using the following targeting techniques were selected based on this principle.

a) T_2-weighted MRI-based targeting: Because of the presence of a high concentration of iron within the STN, which is located lateral to the red nucleus and dorsolateral to the substantia nigra, the anterior part of the nucleus has hypointense signal intensity on T_2-weighted MR images. Although this signal is not always present in the posterior part of the nucleus, it is visible in the anterior portion and correlates well with the red nucleus and substantia nigra to assist the identification of the surgical target location within the STN. The T_2-weighted coronal and axial acquisitions of each patient are fused with his/her pre-operative T_1-weighted image. The target coordinates were defined at a point dorsolateral to the centre of the STN on the three intersecting slices of these overlapped images.

b) Anatomical brain atlas-based targeting: The digitized version of the well-established Schaltenbrand and Wahren stereotactic brain atlas [3] and the segmented STN derived from this atlas were non-rigidly transformed to each patient brain image space using the AtamaiWarp algorithm. A colour-coded atlas, combined with a mesh-representation of the segmented nucleus and the automatically computed STN centroid, permits intuitive planning of the surgical trajectory inside of the nucleus.

Table 2. Imaging parameters of DESPOT1 and DESPOT2

Imaging Modality	TR (ms)	TE (ms)	Flip Angle	BW (kHz)	Voxel Size	Sequence
DESPOT1	11.4	2.9	4°,16°	±7.81	0.34 mm^3	SPGR
DESPOT2	4.2	2.1	15°,55°	±62.5	0.34 mm^3	SSFP

c) T_1 and T_2 map-based targeting: To provide a noninvasive quantitative imaging component to our system, high resolution (0.34 mm^3 isotropic) and signal to noise ratio T_1 and T_2 maps of the deep brain region of a young healthy male volunteer were generated by mutually co-registering and averaging 55 T_1 maps and 25 T_2 maps acquired using the DESPOT1 and DESPOT2 methods respectively [6]. Differences among some deep-brain nuclei that appear indistinct on regular MR images are visible on the two maps. The STN on these maps was segmented using a genetic algorithm (GA) framework and incorporating characteristics of the *k*-means clustering algorithm [6]. Our surgical target identification methodology using the T_1 and T_2 maps is similar to that employed when using the anatomical brain atlas.

d) Electrophysiological database-based targeting: Our electrophysiological database for STN DBS contains intra-operative micro-recording and macro-stimulation data acquired from 49 procedures in standard brain space. The data coded with respect to STN-specific firing patterns were retrieved from the database and

non-rigidly mapped to each patient's MR image to approximately delineate the patient specific nucleus. The location that, when stimulated, resulted in the optimum control of the patient's symptoms was used to define the surgical target.

e) Targeting based on Actual surgical target collection: We non-rigidly registered the collection of final surgical target locations of these 49 STN DBS procedures from the standard brain coordinates to the native brain space of each patient. The automatically generated center of mass (COM) of this cluster of actual targets was directly employed as the target estimation in this targeting protocol.

f) Comprehensive functional and anatomical information-based targeting: With our visualization and navigation system, we localized the surgical targets of these 20 procedures utilizing all the customized functional and anatomical data. The combination of the brain atlas, T_1 and T_2 maps, neurophysiological information, and real surgical data, helped us establish the correlation between the deep-brain anatomical structures and the electrophysiological organization.

3 Results

We computed the mean distances between the actual surgical targets, determined by a stereotactic neurosurgeon using pre-operative MR image-based planning and intra-operative electrophysiological explorations, with those, planned by a non-neurosurgeon using six targeting techniques described above. The results are shown in table 3 below. A one-way analysis of variance (ANOVA) test analyzing the differences on surgical targeting accuracy among the six target localization methods revealed significant differences between groups ($F = 4.48$, $df = 5$, $p = 0.001$). The Tukey's HSD Post-Hoc test also demonstrated that the localization accuracy of comprehensive targeting approach is significantly better than that of the methods based on T_1 and T_2 maps, brain atlas and T_2-weighted MRI ($p < 0.01$). Although no significant difference between the comprehensive approach and the method based on either the electrophysiological database ($p = 0.07$) or the collection of previous final surgical targets ($p = 0.25$), the mean error of the comprehensive targeting approach is still approximately 0.9mm less than that of the two methods. Figure 1 shows that the hypointense signal of STN on the patient T_2-weighted image has better correlation with the segmented nucleus derived from the T_1 and T_2 maps than with that from digitized Schaltenbrand atlas. Figure 2 displays the spatial relationship between the actual surgical target of a patient and those identified with the six targeting methods.

Table 3. Absolute differences between the real surgical targets and the targets estimated with each targeting method

	Mean (mm)	Max (mm)	Min (mm)	Sd (mm)
T_2-weighted MRI	3.0	5.5	2.0	1.3
Brain atlas	3.2	6.2	2.1	1.1
T_1 and T_2 maps	2.9	5.7	2.2	1.1
EP database	2.7	5.8	1.8	1.2
Previous targets	2.5	4.2	1.4	1.0
Combination	1.7	3.1	0.8	0.8

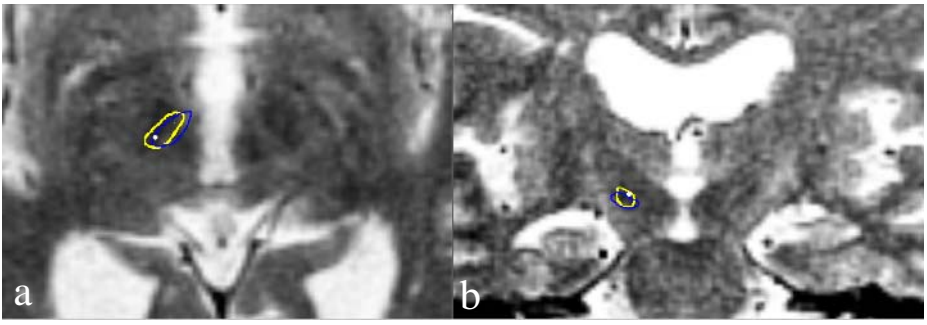


Fig. 1. T_2-weighted MR image of a patient demonstrating direct visualization of STN (a. axial slice, b. coronal slice); Yellow circle: the contour of segmented STN based on T_1 and T_2 maps; Blue circle: based on digitized Schaltenbrand atlas; White sphere: the actual surgical target. The digitized atlas and T_1 and T_2 maps were registered to the patient brain image space.

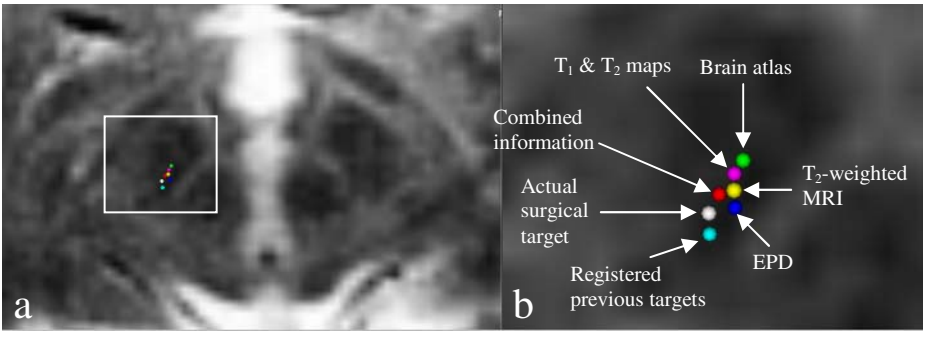


Fig. 2. Positions of surgical targets estimated with each targeting approach. Yellow sphere: target planned based on T2-weighted MR image; Green sphere: based on Schaltenbrand atlas; Magenta sphere: based on T1 and T2 maps; Blue sphere: based on electrophysiological database; Cyan sphere: based on previous surgical targets; Red sphere: based on combination of functional and anatomical data; White sphere: "gold standard" actual surgical target.

In summary, the electrophysiological database, actual surgical target collection, and the integration of approaches based on functional and anatomical information, provided more accurate initial estimation of the surgical target positions than those techniques dependent solely on anatomical references. The location of the surgical target estimated with the comprehensive functional and anatomical information-based targeting method fall within 2 mm of the optimal surgical target, indicating the importance of correlating functional organization with deep-brain anatomy. Practically, the current spread generated at the commonly administered stimulation levels is in the range of 2 mm [14], hence the optimal surgical target region may be easily approached by slightly adjusting the electrode placed at this technique planned initial position. The standard deviation of the combined method was also smaller than that of the other five methods, demonstrating that the targeting precision of this

technique was superior to others. The integration of different functional and anatomical information, contained within the neurosurgical visualization and navigation system, can compensate for their inherent shortcomings and take advantage of their respective strengths to effectively localize the surgical target for STN DBS procedures. Since non-rigid registration was involved in most of the target localization processes, possible errors may partially come from the registration algorithm, which has a mean registration error of 1.0 ± 0.7mm [13].

4 Discussion

In this work, we conducted a retrospective study on 10 patients who were treated with bilateral STN DBS (20 procedures in total) to evaluate the target localization accuracy and precision of six targeting methods in STN DBS surgery. The data of these 10 patients were selected from the database containing data of 26 patients undergone a total of 49 STN DBS procedures at LHSC. The results demonstrated that each method was effective, but the target location estimated with the comprehensive functional and anatomical information-base targeting method was much closer to the optimal surgical target position than that planned using other five approaches. Also, the reduced standard deviation using this targeting method helps ensure precision of localizing the optimal surgical target. Since each of the five methods provides different unique anatomical or functional information, the combination of these references can complement the information provided by the individual methods. Although in some cases the surgical targets defined using the combination of two or three methods are close to those determined by the combination of five, the results vary from case to case and the optimal estimation of the surgical target can not be guaranteed. Using the average position of a repeatedly localized target may reduce the variance and improve the precision of targeting with each method. This can also be achieved by utilizing the comprehensive multiple functional and anatomical references. Because the functional activity does not always exactly correspond to the anatomy, establishing the correlation between the customized population-based electrophysiological database and the patient-specific detailed anatomical structures is necessary to improve the accuracy and precision of targeting. Although accurate surgical target initiation may reduce the need for invasive intra-operative exploration, and thereby decrease the surgical duration and procedure-related complications, certain intra-operative electrophysiological measurements may still be required to compensate possible inadequacy of these targeting methods. During the operation, brain tissues may shift slightly from the pre-operative MR image due to the cranial opening, the electrode insertion, and the cerebrospinal fluid (CSF) leaking. Ideally, one would characterize the degree of brain shift by registering the intra-operative MR image to the pre-operative image. However, due to the unavailability of an intra-operative MR image, the study was conducted by registering the post-operative MR image (in which the electrode placed within the deep brain is displayed) to the pre-operative image using the non-rigid AtamaiWarp algorithm. The results showed that the displacement between the centre of the electrode after registration is within 0.73mm of the planned surgical target on the pre-operative image, demonstrating negligible error caused by brain shift.

Acknowledgements. The authors acknowledge the financial support from the Canadian Institute of Health Research (CIHR), the Ontario Research & Development Challenge Fund (ORDCF), the Canada Foundation for Innovation (CFI), and the Ontario Innovation Trust (OIT).

References

1. C Hamani, E Richter, JM Schwalb, and AM Lozano, "Bilateral subthalamic nucleus stimulation for Parkinson's disease: a systematic review of the clinical literature," *Neurosurgery*, vol. 56, pp. 1313-1324, 2005.
2. EO Richter, T Hoque, W Halliday, AM Lozano, and JA Saint-Cyr, "Determining the position and size of the subthalamic nucleus based on magnetic resonance imaging results in patients with advanced Parkinson disease," *J. Neurosurgery*, vol. 100, pp. 541-546, 2004.
3. G Schaltenbrand and W Wahren, *Atlas for Stereotaxy of the Human Brain*. Stuttgart, Germany: Thieme, 1977.
4. KA Ganser, H Dickhaus, R Metzner, and CR Wirtz, "A deformable digital brain atlas system according to Talairach and Tournoux," *Med. Imag. Analy*., vol. 8, pp. 3-22, 2004.
5. KV Slavin, KR Thulborn, C Wess, and H Nersesyan, "Direct visualization of the human subthalamic nucleus with 3T MR imaging," *Am. J. Neuroradiol*., vol. 27(1), pp. 80-84, 2006.
6. SCL Deoni, BK Rutt, and TM Peters, "Segmentation of thalamic nuclei using high resolution quantitative magnetic resonance imaging and a modified k-means algorithm," in *Proc. 13th ISMRM*, no. 168, 2005.
7. KW Finnis, YP Starreveld, AG Parrent, AF Sadikot, and TM Peters, "Three-dimensional database of subcortical electrophysiology for image-guided stereotactic functional neurosurgery," *IEEE Trans. Med. Imag*., vol. 22(1), pp. 93–104, 2003.
8. PF D'Haese, E Cetinkaya, PE Konrad, C Kao, and BM Dawant, "Computer-aided placement of deep brain stimulators: from planning to intraoperative guidance," *IEEE Trans. Med. Imag*., vol. 24(11), pp. 1469-1478, 2005.
9. T Guo, KW Finnis, AG Parrent, and TM Peters, "Development and application of functional databases for planning deep-brain neurosurgical procedures," in *Proc. MICCAI 2005,* J Duncan and G Gerig (Eds.), LNCS 8749, pp. 835- 842, 2005.
10. E Cuny, D Guehl, P Burbaud, C Gross, V Dousset, and A Rougier, "Lack of agreement between direct magnetic resonance imaging and statistical determination of a subthalamic target: the role of electrophysiological guidance," *J. Neurosurg*., vol. 97, pp. 591-7, 2002.
11. C Hamani, EO Richter, Y Andrade-Souza, W Hutchison, JA Saint-Cyr, AM Lozano, "Correspondence of microelectrode mapping with magnetic resonance imaging for subthalamic nucleus procedures," *Surg. Neurol.*, vol. 63(3), pp. 249-253, 2005.
12. YP Starreveld, "Fast nonlinear registration applied to stereotactic functional neurosurgery," Ph.D. dissertation, Univ. Western Ontario, London, ON, Canada, 2002.
13. T Guo, YP Starreveld, and TM Peters, "Evaluation and validation methods for intersubject non-rigid 3D image registration of the human brain," in *Proc. SPIE Medical Imaging*. Bellingham, WA: SPIE, 2005, vol. 5744, pp. 594-603, Visualization, Image-Guided Procedures, and Display.
14. KA Follett, MD Mann, "Effective stimulation distance for current from macroelectrodes," *Exp. Neurol.*, vol. 92(1), pp. 75-91, 1986.

Objective Outcome Evaluation of Breast Surgery

Giovanni Maria Farinella[1], Gaetano Impoco[1], Giovanni Gallo[1], Salvatore Spoto[1], Giuseppe Catanuto[2], and Maurizio B. Nava[2]

[1] IPLab - DMI , University of Catania, Italy
{gfarinella, impoco, gallo}@dmi.unict.it
[2] Unit of Plastic and Reconstructive Surgery - INT, Milano, Italy

Abstract. A new method is proposed to unambiguously define a geometric partitioning of 3D models of female thorax. A breast partitioning scheme is derived from simple geometric primitives and well-defined anatomical points. Relevant measurements can be extrapolated from breast partition. Our method has been tested on a number of breast 3D models acquired by means of a commercial scanner on real clinical cases.

1 Introduction and Motivation

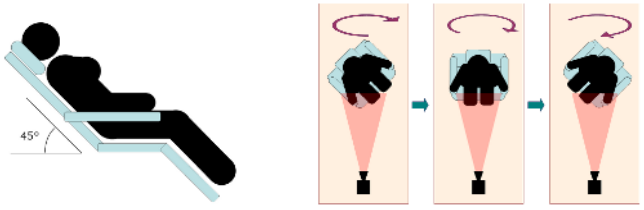
Analysis and definition of breast shape based on clinical semantics can help plastic surgeons to objectively evaluate the outcome of plastic surgery. Until today, visual assessment is the common clinical practice. Although several reproducible measurement methodologies have been proposed in the literature, none of them reached clinical practice. Some authors propose to measure the breast volume by means of water displacement, MRI, X-rays, or thermoplastic moulding [1,2,3,4,5,6]. Others use anatomical landmarks and manual measurements. However, their measurement techniques are inaccurate (e.g., area of breast surface, length of the path over the breast surface, and so on). Moreover, manual measurements heavily depend on experimental settings, such as the posture of the patient. 3D scanning technologies can be exploited to acquire useful data for accurate and reproducible analysis of breast surface.

We define a non-ambiguous geometric segmentation of 3D models of the thoraco-mammary surface, starting from the informal specifications of the surgeons. We also indicate a set of measurements and visualisation methods to objectively evaluate the outcome of breast surgery, and to provide a sound and objective meaning to vague terms, such as *sweet slope* and *well defined inferior pole*. The proposed partitioning scheme has been implemented in a custom tool named Breast Shape Analyzer 0.1 (BSA 0.1).

2 Breast Surface Analysis

The models used in our work were collected according to common cosmetic and reconstructive criteria (e.g., age, menopausal status, ptosis classification). A commercial laser scanner has been used to acquire the data.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 776–783, 2006.



(a) Patients' posture at 45 degrees.

(b) Rotation of the chair with respect to the camera.

Fig. 1. Patient positioning

The recruited volunteers were seated on a chair with their back at 45 degrees (Figure 1). Three scans of the same volunteer were acquired: facing the camera and rotating the chair at 45 degrees to the left and to the right. We choose not to merge the different scans because of uncontrollable physiological movements (e.g., breathing and imperceptible body torsions) during the scanning process. The time lag required by laser scanning is not compatible with the patient breathing. This difficulty may be overcome by adopting faster but less accurate shape-from-video techniques [7].

Therefore, the analysis has been performed on partial views. Unsampled areas have been filled using volumetric diffusion [3]. This method is compatible with precision requirements of surgeons when it deals with small gaps (below 0.8 cm of diameter). Gaps of larger size have been filled only for visualisation purposes and the reconstructed areas have not been used in landmark positioning.

2.1 Anatomical Landmarks

In order to unambiguously define breast sub-units, we exploited the following anatomical landmarks, suggested by surgeons because of their easy reproducibility (Figure 2):

1. *Sternal Notch* or *Jugulum* (p_j)
2. *Xiphoid* (p_x)
3. *Nipple* (p_c)
4. *Pectoralis insertion in the arm* (p_{aa})
5. *Acromial extremity of clavicle* (p_s)
6. *Mid-axillary point* (p_{pa})
7. *Lowest breast point with respect to the vertical body axis* (p_d)

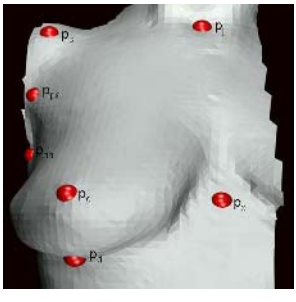


Fig. 2. Anatomical landmarks

These points are interactively placed by surgeons on the model using the BSA 0.1 tool described in Section 3.

2.2 Anatomical Sub-units

In clinical practice, breast sub-units are manually traced and the outcome strongly depends on the experience and ability of surgeons. We aimed at defining a tracing scheme that is less dependent on the ability of the user. The proposed scheme employs simple geometric primitives such as planes, lines, and geodesics. Geometrically-defined breast partitioning guarantees a more objective and reproducible subdivision procedure and opens the way to objective quantitative measuring methods. The proposed subdivision algorithm defines four elements:

1. *Thoraco-mammary bilateral symmetry plane* (π_s);
2. *Breast meridian plane* ($\pi_{m,l}$ and $\pi_{m,r}$ for left and right breast, respectively);
3. *Breast equatorial plane* ($\pi_{e,l}$ and $\pi_{e,r}$ for left and right breast, respectively);
4. *Breast box hull.*

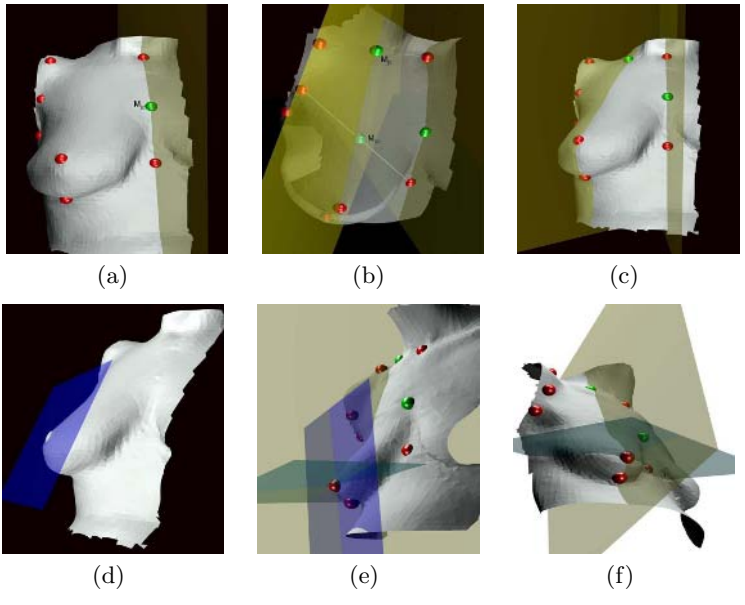


Fig. 3. Planes used for breast segmentation. (a) Thoraco-mammary bilateral symmetry plane. (b) Breast meridian plane (c) Breast meridian curve obtained by intersecting the 3D model and the meridian plane. (d)Nipple-areola plane. (e) Construction of the equatorial plane. (f) Breast equatorial curve.

First, we define the *bilateral symmetry plane* as the axial plane through p_j (*jugulum*), p_x (*xiphoid*) and the derived point M_{jx}. M_{jx} is the middle point of the geodesic between p_j and p_x. Notice that these three points are not collinear due to the natural curvature of the thorax (Figure 3(a)).

The *breast meridian plane* is defined by three points (see Figure 3(b)): the nipple, the midpoint, M_{js}, of the straight line between p_j (sternal notch) and p_s (acromial extremity of the clavicle), and the midpoint M_{px} of the straight line between p_{pa}

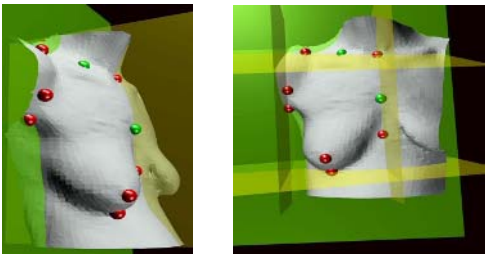


Fig. 4. Construction of the *Breast Box Hull*

(pectoralis insertion in the arm) and p_x (xiphoid). This plane intersects the model subdividing the surface in lateral and medial aspects. The curve produced on the model is usually referred to as *breast meridian* (Figure 3(b) and Figure 3(c)).

In order to define the *equatorial plane*, we imitate a very common device used by plastic surgeons to mark the nipple-areola complex. The *nipple-areola plane* is defined as the tangent plane at the nipple, p_c (Figure 3(d)). The equatorial plane is orthogonal to the nipple-areola plane and to the line defined by the intersection of this plane and the meridian plane (Figure 3(e)). The intersection between the equatorial plane and the breast surface is called *breast equator*. It separates the upper and lower halves of the breast (Figure 3(f)).

A region of interest (*breast box hull*) has been selected in compliance with surgeons' specifications to analyse breast surface properties. It is made up of six surfaces:

- *Breast surface*
- *Back plane*, π_b, orthogonal to π_s and containing p_s and p_{aa} Figure 4(a))
- *Thoraco-mammary bilateral symmetry plane* π_s (Figure 3(a))
- *Lateral plane*, $\pi_{l,i}$, orthogonal to π_b and containing p_{aa} and p_{pa} (Figure 4(b))
- *Lower plane*, π_L, orthogonal to π_s and passing through p_d (Figure 4(b))
- *Upper plane*, π_U, orthogonal to π_s and to π_b and including p_j and p_s (Figure 4(b))

2.3 Measurements

Given the partition defined in Section 2.2 a number of relevant measurements can be computed:

Line Measurements. Distances between all the landmark pairs and distances between p_c and the bilateral symmetry plane, and between the nipple, p_c, and the back plane, π_b (Figure 5).

Area measurements. We computed the area of some relevant surfaces: the front box hull, the upper pole, the lower pole, and the sub-units defined by the intersection of the meridian plane, $\pi_{m,i}$, and the equator plane, $\pi_{e,i}$.

Angles. Let us define $O_i x_i y_i z_i$ the (generally non-orthogonal) coordinate system induced by the intersection of the planes π_s, $\pi_{m,i}$, $\pi_{e,i}$ (refer to Section 2.2 and Figure 6), where i stands for l (left) or r (right). We compute

the angles $\widehat{x_i y_i}$ and $\widehat{x_i z_i}$, where x_i is the axis including the nipple, p_c. The former gives a measure of the *divergence* of the nipples with respect to the bilateral symmetry plane. The latter quantifies the breast*ptosis*.

Curvature and convexity. Visualising the mean curvature and convexity of the breast surface is useful to capture the concept of widely used clinical terms, such as "sweet slope" and "well defined inferior pole". Since no unambiguous definition of breast shape features is available, surgeons found false-colour visualisation more useful than any attempt to make accurate measurements of data.

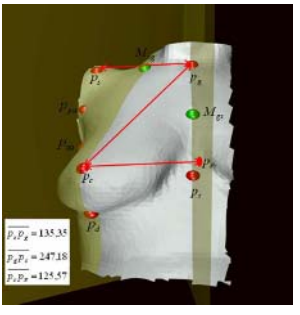


Fig. 5. Line measurements

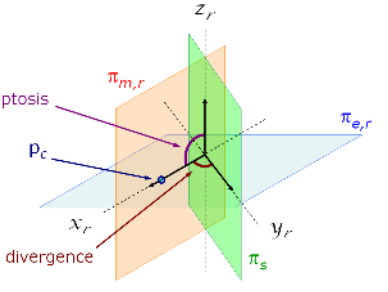
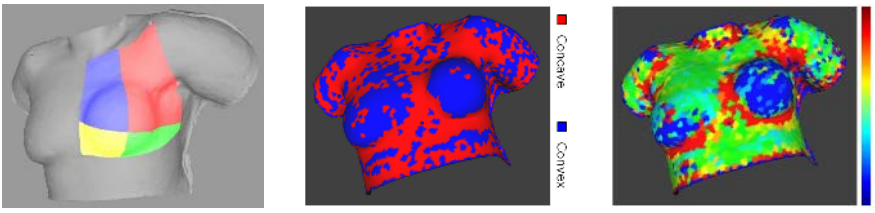


Fig. 6. The Angle Reference System

3 The *Breast Shape Analyzer* Tool

The Breast Shape Anayzer tool (BSA 0.1) allows the positioning of landmarks directly on the 3D model. When a point in screen space is acquired, a ray is cast from the near plane to the far plane, in the OpenGL context, and the closest vertex to this ray is automatically selected as a landmark point. Every landmark shift on the mesh is interactively shown and the coordinates values are visualised in a separate text area of the interface for feedback. The tool automatically computes and shows all the relevant planes as soon as all the needed landmarks are positioned. A critical issue is the computation of the tangent plane at the nipple, p_c. This is done using an approximation of the normal to the surface. The segmented subunits may be visualised in different contrasted colours (Figure 7(a)). All the measures that have been listed in the previous section can be calculated, visualised, and stored.

Finally, the model is painted in false colours according to the mean curvature and to the convexity of the breast surface. Two colouring modes are allowed: curvature and convexity (Figures 7(b) and 7(c)). Curvature is displayed using a colour ramp ranging from blue to red, in increasing order of curvature. Convexity is encoded using two colours, representing concave and convex regions. Surface colouring significantly captures breast shape features, thus guiding surgeons in post-operative outcome evaluation. Colouring does not fulfill our expectations due to the sampling noise of the surface. This effect can be partially reduced by smoothing the surface, while preserving significant details (i.e., high curvature regions). Edge-preserving



(a) Breast subunits. (b) Concavity / convexity. (c) Curvature.

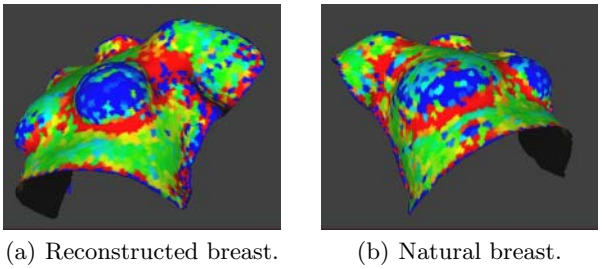
Fig. 7. Visualisation of the 3D model of a volunteer's breast using the BSA tool

smoothing can be implemented using many different filters, such as Bilateral filter [8]. Alternatively, curvature can be computed on a simplified mesh and then mapped back to the original model e.g., using a point-surface distance.

4 Clinical Test

Most of the published medical studies on breast shape evaluation focus on reconstruction using volume estimation or simple linear measurements [1,2,3,4,5,6]. None of them has reached clinical practice since either the landmarks are ambiguously and arbitrarily placed, or their definition is hard to reproduce in complex cases. We place landmarks on bony extremities, since they are more reproducible. Building on these points, the resulting unambiguous geometric partitioning of the breast shape is useful to develop a reliable breast topography and a proper outcome evaluation protocol. Our method cannot be compared to any other existing technique due to lack of objective methods for breast shape measurement in the literature as well as in clinical practice. We rather demonstrate the capabilities of our method to capture commonly-used clinical evaluation parameters by analysing a case study.

In Table 1 and Figure 8, we report a clinical case of a woman (Patient CH) who underwent mastectomy and reconstruction with permanent silicone implants and developed a Baker Grade IV capsular contracture. Our measurements correctly predict a severe alteration of biometric parameters, such as shoulder *adduction*



(a) Reconstructed breast. (b) Natural breast.

Fig. 8. Local mean curvature

Table 1. Relevant measurements for patient Ch

		Reconstructed breast	Natural breast
Surfaces Area (cm^2)	Total	461.00	522.30
	Upper outer quadrant (Red)	147.30	218.00
	Upper inner quadrant (Blue)	60.38	74.50
	Lower outer quadrant (Green)	187.00	171.20
	Lower inner quadrant (Yellow)	66.00	57.00
Upper/lower surface ratio		0.82	1.28
Distances (cm)	$\overline{p_c \pi_s}$	10.00	11.10
	$\overline{p_c p_j}$	18.40	20.20
	$\overline{p_c p_s}$	18.90	19.40
	$\overline{p_c p_{aa}}$	14.50	15.80
	$\overline{p_j p_s}$	13.40	14.30
	$\overline{p_j p_{aa}}$	14.50	15.82
	$\overline{p_j p_{aa}}$	15.20	16.70
	$\overline{p_x p_{aa}}$	21.90	23.20
Angles (degrees)	Ptosis	95	84
	Divergence	39	31

($\overline{p_j p_s} = 13.4cm$ on the reconstructed side and $14.3cm$ on the natural side) and *flexion* of vertebral spine ($\overline{p_x p_{aa}} = 21.9cm$ on the reconstructed side and $23.2cm$ on the natural side). The reconstructed nipple lies $18.4cm$ apart from the jugulum, while this distance is $20.2cm$ on the natural side. Surface area measurements show an overall reduction of the breast box hull surface after the operation (461 cm^2 reconstructed, 522 cm^2 natural). Moreover, the area ratio between the upper and the lower halves of the breasts show a predominance of the inferior half of the reconstructed side, while on the natural side the upper areas are wider. We also introduce the *divergence angle* (see Section 2.3) as a new parameter for breast shape evaluation. This measurement is considered by surgeons capable of capturing important features of the breast shape. Basically, it expresses the patient's observation point on her breast (from the top), a feature that cannot be guessed by doctors staying in front of her. The analysis of cases who underwent bilateral augmentation, and whose breasts are modelled by two identical implants underneath, show similar estimates on both sides. Finally, the false-colour visualisation (Figure 8) of the local mean curvature of the breast surface is considered particularly useful by surgeons. The figure shows the alteration of a reconstructed breast in the clinical case CH. Highly curved surfaces are predominant on the reconstructed side, while the natural breast is smother and almost flat close to the upper-inner quadrants. Although this information can be collected from visual in *vivo* observation or from observation of 2D photographs, these procedures do not allow objective evaluation and data exchange between surgeons.

The reproducibility of the reported numerical measures is assured by the partitioning algorithm defined in Section 2.2. We are currently evaluating the effect of landmark misplacement of the final result.

5 Conclusions

Evaluating the post-operative outcome of breast plastic surgery is currently a little more than craftsmanship and intuition, strongly relying on the professional skills of surgeons. Clinical capabilities and personal experience of plastic surgeons

cannot be embedded into a standard clinical procedure, unless a new theoretical background in anatomical shape analysis is established.

A series of feature points uniquely characterising woman breast surface was fixed in compliance with surgeon's suggestions and requirements. Building on these landmarks, we devised a simple and reproducible breast partitioning scheme employing only geometric primitives as simple as planes, lines, and geodesics. We came out with a simple and intuitive description of breast surface that has been appreciated by the plastic surgery community and is being clinically validated. Our unambiguous geometric breast partitioning is a first step towards the definition of a portable and reproducible analysis scheme, and opens the path to the definition of objective breast biometrics that, in turn, can lead to the definition of a standard clinical evaluation protocol.

A prototype software, *Breast Shape Analyzer* (BSA 0.1), based on our novel methodology, has been produced to empower surgeons with useful measuring and visualisation tools. BSA will be soon available in our website . Many challenging issues are left open by this work:

a) use of faster (and/or cheaper) scanning techniques;
b) modelling of the elastic and structural deformation of the breast shape;
c) parametric fitting of a standard breast model from few 3D anatomical landmarks and/or a low resolution scan of the body.

Future works will be focussed on these issues.

References

1. Brown, T.P., Ringrose, C., Hyland, R.E., Cole, A.A., M., B.T.: A method of assessing female breast morphometry and its clinical application. British Journal of Plastic Surgery **52**(5) (1999) 355–359
2. Tebbetts, J.B.: A system for breast implant selection based on patient tissue characteristics and implant-soft tissue dynamics. (2002)
3. Davis, J., Marshner, S., Garr, M., Levoy, M.: Filling holes in complex surfaces using volumetric diffusion. In: First Int. Symp. on 3D Data Processing, Visualization and Transmission (3DPVT'02), IEEE Comp. Soc. (2002) 428–438
4. Galdino, G.M., Nahabedian, M., Chiaramonte M., Geng, J.Z.K.S., Manson, P.: Clinical applications of three-dimensional photography in breast surgery. Plastic and Reconstructive Surgery **110**(1) (2002) 58–70
5. Lee, H.Y., Hong, K. Kim, E.A.: Measurement protocol of women's nude breasts using a 3d scanning technique. **35**(4) (2004) 353–360
6. Losken, A., Seify, H., Denson, D.D., Paredes, A.A.J., Carlson, G.W.: Validating three-dimensional imaging of the breast. **55**(6) (2005) 695–696
7. Pollefeys, M.: 3D from image sequences: Calibration, motion and shape recovery. In Paragios, N., Chen, Y., Faugeras, O., eds.: Mathematical Models in Computer Vision: The Handbook. Springer (2005)
8. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: ICCV '98: Proceedings of the Sixth International Conference on Computer Vision, Washington, DC, USA, IEEE Computer Society (1998)

Automatic Detection and Segmentation of Ground Glass Opacity Nodules

Jinghao Zhou[1], Sukmoon Chang[1,2], Dimitris N. Metaxas[1], Binsheng Zhao[3], Lawrence H. Schwartz[3], and Michelle S. Ginsberg[3]

[1] CBIM, Rutgers, The State University of New Jersey, NJ, USA
jhzhou@eden.rutgers.edu, dnm@cs.rutgers.edu
[2] Computer Science, Capital College, Penn State University, Middletown, USA
sukmoon@psu.edu
[3] Department of Radiology, Memorial Sloan-Kettering Cancer Center, NY, USA
zhaob@mskcc.org, schwartl@mskcc.org, ginsberm@mskcc.org

Abstract. Ground Glass Opacity (GGO) is defined as hazy increased attenuation within a lung that is not associated with obscured underlying vessels. Since pure (nonsolid) or mixed (partially solid) GGO at the thin-section CT are more likely to be malignant than those with solid opacity, early detection and treatment of GGO can improve a prognosis of lung cancer. However, due to indistinct boundaries and inter- or intra-observer variation, consistent manual detection and segmentation of GGO have proved to be problematic. In this paper, we propose a novel method for automatic detection and segmentation of GGO from chest CT images. For GGO detection, we develop a classifier by boosting k-NN whose distance measure is the Euclidean distance between the nonparametric density estimates of two examples. The detected GGO region is then automatically segmented by analyzing the texture likelihood map of the region. We applied our method to clinical chest CT volumes containing 10 GGO nodules. The proposed method detected all of the 10 nodules with only one false positive nodule. We also present the statistical validation of the proposed classifier for GGO detection as well as very promising results for automatic GGO segmentation. The proposed method provides a new powerful tool for automatic detection as well as accurate and reproducible segmentation of GGO.

1 Introduction

Ground Glass Opacity (GGO) is defined as hazy increased attenuation within a lung that is not associated with obscured underlying vessels, but with preservation of bronchial and vascular margins [1]. It can reflect minimal thickening of the septal or alveolar interstitium, thickening of alveolar walls, or the presence of cells or fluid filling the alveolar spaces. It can represent active disease such as pulmonary edema, pneumonia, or diffuse alveolar damage. The results of the Early Lung Cancer Action Project, or ELCAP, suggested that nodules with pure (nonsolid) or mixed (partially solid) GGO at the thin-section CT are more likely to be malignant than are those with solid opacity [2]. A focal area of

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 784–791, 2006.

pure GGO on the thin-section CT seems to be an early sign of bronchioloalveolar carcinoma (BAC) [3]. Pure GGO is useful for differentiating small localized BAC from small adenocarcinomas not having a replacement growth pattern [4]. Detection and treatment of pure GGO can improve a prognosis of lung cancer [5].

The appearances of GGO on CT images such as its shape, pattern, and boundary are very different from solid nodules. Thus, algorithms developed for segmentation of solid nodules are very likely to produce inaccurate results when applied to GGO. In [6], a hybrid neural network of three single nets and an expert rule are applied to detect GGO. This method underestimates GGO area due to its improper cut-off of the edges of GGO. Hence, this method may be used only for large GGO and may not be able to obtain accurate segmentation for small GGO. [7] detected GGO using automatic clustering techniques and focused only on GGO detection. The GGO segmentation was not discussed in their work. [8] applies GGO segmentation method based on Markov random field and vessel removal method based on shape analysis. However, they only focused on GGO segmentation. The GGO detection was not discussed in their work.

In this paper, we propose a novel method for automatic detection and segmentation of GGO from chest CT images. For GGO detection, we develop a classifier by boosting k-NN whose distance measure is the Euclidean distance between the nonparametric density estimates of two examples. The detected GGO region is then automatically segmented by analyzing the texture likelihood map of the region. We also present the statistical validation of the proposed classifier for GGO detection as well as the very promising results of automatic GGO segmentation. The proposed method provides a new powerful tool for automatic detection as well as accurate and reproducible segmentation of GGO.

2 Method

2.1 GGO Detection

We first develop a novel method for automatic GGO detection from chest CT images. The method is based on texture analysis using a machine learning framework, i.e., boosting the k-NN classifier. Our primary focus in this section is the accurate classification of GGO from other objects in chest CT images.

Vessel and noise suppression with Cylinder filters. The accuracy of GGO detection may be hindered by various structures within a lung. To avoid this difficulty, we first develop a cylinder filter to suppress intensity values of vessels and other elongated structures as well as noise inside a lung, while maintaining GGO intensity values intact [9]. The cylinder filter F_{cyl} is defined as:

$$F_{\text{cyl}}(x) = \min_{\theta} \left(\min_{y \in \Omega_{\theta}^{x}} I(y) \right)$$

where, Ω_{θ}^{x} is the domain of the cylinder filter centered at x with orientation θ. F_{cyl} is a hybrid neighborhood proximity filter that produces strong responses to

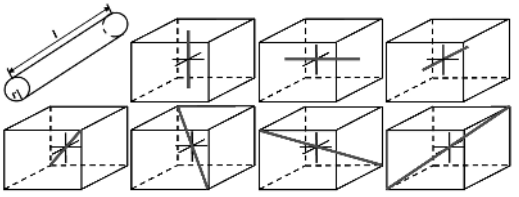


Fig. 1. Cylinder filter and its 7 orientations

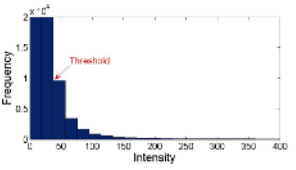
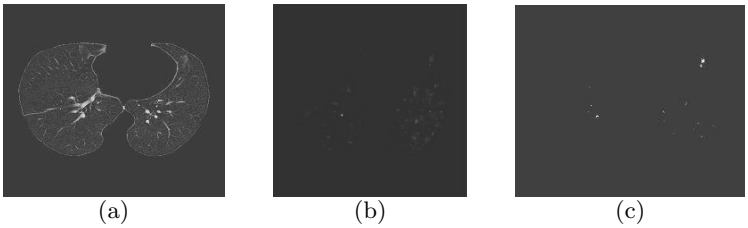


Fig. 2. Histogram of a cylinder filtered image



(a) (b) (c)

Fig. 3. Effect of F_{cyl}. (a) Original image (b) Image filtered with F_{cyl} (c) After thresholding.

blob-like objects (i.e. GGO). In this paper, we have selected the parameters of F_{cyl} empirically and used a cylinder with radius of 1, 2 and 3 voxels and length of 7 voxels at 7 different orientations, as shown in Fig. 1. The effect of F_{cyl} is shown in Fig. 3. In the figure, (b) shows the filter-response image of F_{cyl} applied to (a). We can see from (b) that vessels and noise are effectively suppressed while GGO remains intact.

To isolate candidate GGO areas, i.e., areas of high response values, we apply a simple thresholding to the filter-response image (Fig. 3(c)). The threshold value is automatically determined by analyzing the histogram of the filter-response image, as illustrated in Fig. 2. In the next section, we apply a machine learning framework to classify the candidate GGO areas.

Classification of GGO by boosting *k*-NN. k-NN assumes that an instance is represented as a point in n-dimensional space, $\mathcal{R}^n$. The nearest neighbors of an instance are defined in terms of the standard Euclidean distance. An instance, x, can be denoted by a feature vector, $\langle a_1(x), a_2(x), \ldots, a_n(x)\rangle$, where $a_r(x)$ denotes the value of the r^{th} attribute of x. Then, the distance between two instances x_i and x_j is defined as:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{n} (a_r(x_i) - a_r(x_j))^2}$$

Fig. 4 shows the typical intensity pdf (probability density function) of positive and negative examples. We can see from the figure that positive and negative

examples are well separable using the pdf. For this reason, we use as an instance the intensity pdf of the region surrounding a candidate GGO.

The learning for the classification of the possible GGO areas isolated from the previous section has a discrete target function of the form $f : \mathcal{R}^n \mapsto \mathcal{V}$, where $\mathcal{V} = \{\oplus, \ominus\}$, with the label $\oplus$ for GGO and $\ominus$ for non GGO areas. The k-NN algorithm returns $\hat{f}(x_q)$ as its estimate of $f(x_q)$, which is the most common value of f among the k training examples nearest to x_q:

$$\hat{f}(x_q) = \arg\max_{v \in \mathcal{V}} \sum_{i=1}^{k} \delta(v, f(x_i))$$

where, x_q is a query instance to be classified, $x_1, \ldots, x_k$ denote the k instances from training samples that are nearest to x_q, and $\delta(a, b) = 1$ if $a = b$ and $\delta(a, b) = 0$ otherwise. To obtain an accurate classification, k-NN needs a large training set, which results in slow classification due to the large number of distance calculations. We overcome this difficulty by boosting the k-NN.

Boosting is a method for improving the performance of any weak learning algorithm which, in theory, only needs to perform slightly better than random guessing. A boosting algorithm called **AdaBoost** improves the performance of a given weak learning algorithm by repeatedly running the algorithm on the training data with various distributions and then combining the classifiers generated by the weak learning algorithm into a single final classifier [10]. As in [10], our purpose of using **AdaBoost** is to improve the speed of k-NN by reducing the number of prototype samples and thus reducing the required number of distance calculation without affecting the error rate. The details of boosting k-NN is given in [10]

2.2 GGO Segmentation in ROI

Because of the hazy appearance of GGO and the large overlap of intensity values between GGO and surrounding vessels, simple thresholding or edge based segmentation method do not produce acceptable results for GGO segmentation. The proposed method applies a nonparametric density estimation and likelihood map method based on the texture of GGO for segmentation [11], followed by eigenanalysis of the Hessian matrix to remove vessels overlapped with GGO [12].

Nonparametric density estimation. Let $\mathcal{R}_M$ be the volumetric region of a candidate GGO bounded by a cube model Φ_M. We estimate the probability density function (pdf) of intensity values of the model interior. We use a nonparametric kernel based method to approximate the pdf. Let i, $i = 0, \ldots, 255$, denote the random variable for intensity values. The intensity pdf of the model-interior region is defined by:

$$P(i|\Phi_M) = \frac{1}{V(\mathcal{R}_M)} \iiint_{\mathcal{R}_M} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(i - I(y))^2}{2\sigma^2}\right) dy$$

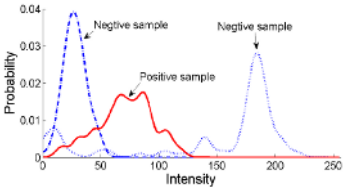


Fig. 4. Typical probability density functions of positive (solid), negative (dash-dot, typical background), and negative (dotted, typical vessels) examples

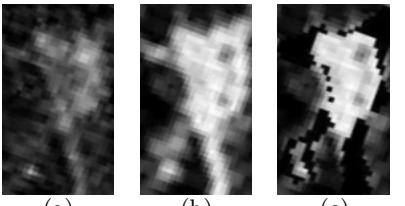


Fig. 5. (a) Original image of GGO, (b) likelihood map in ROI computed based on texon statistics method, and (c) Likelihood map in ROI after vessel removal based on eigenanalysis of the Hessian matrix

where, $V(\mathcal{R}_M)$ denotes the volume of $\mathcal{R}_M$, y are the pixels in the domain $\mathcal{R}_M$, and σ is the standard deviation of a Gaussian kernel.

We use an information theoretic distance measure called Kullback-Leibler Divergence (KLD) to measure the dissimilarity between 2 pdf's described in the previous section [13]. The symmetrized variation of KLD is called Chernoff Information. The Chernoff Information between p_1 and p_2 is defined by:

$$C(p_1 \| p_2) = \max_{0 \le t \le 1} (-\log \mu(t))$$

where, $\mu(t) = \int [p_1(i)]^{1-t} [p_2(i)]^t \, di$. A special case of Chernoff distance is the Bhattacharya distance, in which t is chosen to be $\frac{1}{2}$, i.e., the Bhattacharya distance between p_1 and p_2 is:

$$B(p_1 \| p_2) = -\log \mu(\frac{1}{2}) \tag{1}$$

Also, we denote:

$$\rho(p_1 \| p_2) = \mu(\frac{1}{2}) = \int [p_1(i)]^{\frac{1}{2}} [p_2(i)]^{\frac{1}{2}} \, di \tag{2}$$

Eq. (1) and (2) shows that, when the value of ρ goes from zero to one, the corresponding value of B goes from infinity to zero. The Chernoff Information is an important information-theoretic distance measure, and it has been shown this measure is the exponential rate of optimal classifier performance probabilities [14].

Texture likelihood map. We now evaluate the likelihood of the neighborhood sphere texon of every pixel in ROI. We define this likelihood using ρ, since it increases as the Bhattacharya distance between two distributions decreases. That is, the likelihood of any pixel x belonging to the object of interest is defined by:

$$L(T(x, \hat{s}) | \Phi_M) \propto \rho(p_{T(x,\hat{s})} \| p_m)$$

where, $T(x, \hat{s})$ represents the neighborhood sphere texon centered at x with scale $\hat{s}$, $p_{T(x,\hat{s})}$ is the intensity pdf of the texon, and p_m is the learned pdf of Φ_M. Fig. 5(b) shows the texture likelihood map of the image in (a). The radius of texons in our paper is less than 3 pixels and the model interior texture is mostly homogeneous with some level noise. Thus, it is not necessary to consider the spatial correlation between pixels.

Vessel removal. Finally, we remove the vessels around GGO in the texture likelihood map. The eigenanalysis of the Hessian matrix is a widely used method for vessel detection [12]. Given an image $I(x)$, the local intensity variations in the neighborhood of a point x_0 can be expressed with its Taylor expansion:

$$I(x_0 + h) \approx I(x_0) + h^T \nabla I(x_0) + h^T H(x_0) h$$

where, $\nabla I(x_0)$ and $H(x_0)$ denote the gradient and the Hessian matrix at x_0, respectively. $H(x_0)$, whose components are the second order derivatives, describes the curvature of $I(x)$ at x_0. Let $\lambda_1, \lambda_2, \lambda_3$ and $\boldsymbol{e_1}, \boldsymbol{e_2}, \boldsymbol{e_3}$ be the eigenvalues and eigenvectors of H such that $\lambda_1 \leq \lambda_2 \leq \lambda_3$ and $|\boldsymbol{e_i}| = 1$. The signs and ratios of the eigenvalues provide the indications of various shapes of interest, as summarized in [15]. Fig. 5(c) shows the texture likelihood map after the removal of vessels in (b).

3 Results

To test the GGO detection method, we collected 319 example images, containing 200 training examples and 119 test examples. The example images were of size 9× 9 pixels extracted from the CT volumes. The distance that we used for k-NN to identify the nearest neighbor and to classify an instance is the standard Euclidean distance between the nonparametric density estimates (represented as vectors in $\mathcal{R}^{256}$) of two examples as described in Sec. 2.2. We performed bootstrapping to estimate the generalization error of our GGO detection method [16]. We trained and tested the method on bootstrap samples. After 20 steps of boosting, the test error rate converges to 5.04%.

We applied the trained GGO classifier to 10 CT volumes containing 10 nodules. The CT volumes were acquired by multi-slice HRCT scanners with 1 mm slice collimation. The number of slices in each CT scan ranged from 23 to 29 (interpolated to 92 to 116), each of which consists of a 512 × 512 matrix, with in-plane resolution ranging from 0.57mm to 0.71mm. The classifier detected the total of 11 nodules, containing all of the 10 nodules and one false positive nodule.

The detected nodules were then segmented as described in Sec. 2.2. Fig. 6 shows the GGO segmentation results overlaid on the original CT images and 3D reconstruction for three GGOs as examples. From the figure, we can see that the surrounding vessels are accurately removed from the GGO segmentation. These results demonstrate the potential of our segmentation method to deal with haze patterns of GGO.

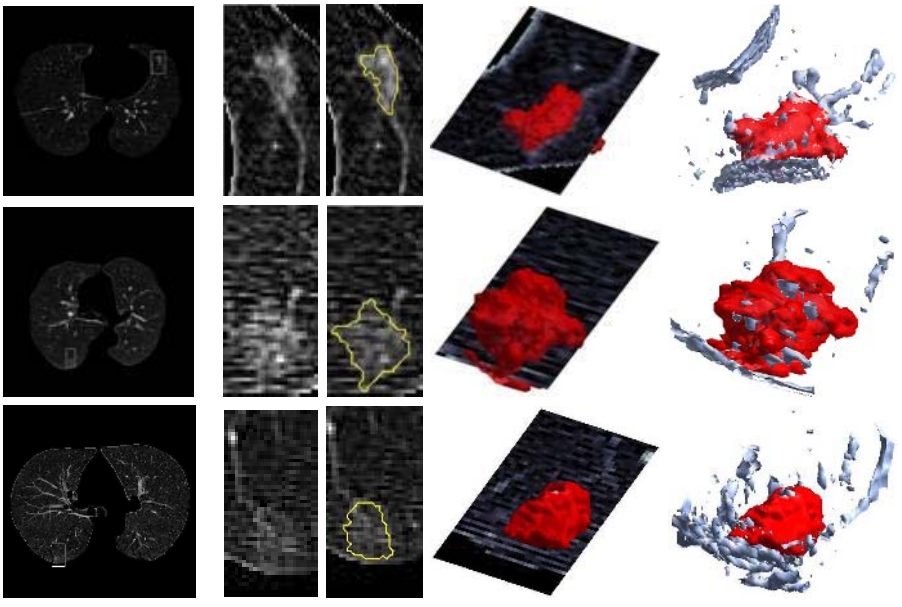


Fig. 6. GGO segmentation. (a) original CT images, (b) enlarged CT images, (c) segmentation overlay, (d)and (e) 3D reconstruction of segmented GGO

4 Conclusion

We proposed a novel method for the automatic detection and segmentation of GGO from chest CT images. The proposed method consists of two steps, i.e., GGO detection, followed by GGO segmentation. For GGO detection, vessels and noise were first suppressed using cylinder filters. Then, candidate GGO regions were extracted by thresholding. We automatically selected the threshold by the histogram analysis of the filter-response images. Finally, the candidate GGO regions were classified by boosting the k-NN, whose distance measure was the Euclidean distance between the intensity pdf's of two examples. The validation of the proposed method using bootstrapping shows the mean error rate of 5.04%. Our method applied to clinical chest CT volumes containing 10 GGO nodules also shows the promissing result, detecting all of the 10 nodules with one false positive nodule.

Each GGO region identified was automatically segmented by analyzing the texture likelihood map of the region. We presented various results of GGO detection and segmentation from clinical chest CT images. The manual segmentation of GGO has proved to be problematic due to large inter-observer variations as well as large intra-observer variations. The proposed method introduces a novel automatic tool for accurate detection as well as reproducible segmentation of GGO.

References

1. Austin, J., Muller, N., Friedman, P.e.a.: Glossary of terms for CT of the lung: recommendations of the nomenclature committee of the fleischner society. Radiology **200** (1996) 327–331
2. Henschke, C., Yankelevitz, D., Mirtcheva, R.: CT screening for lung cancer: frequency and significance of part-solid and nonsolid nodules. Am. J. Roentgenol. **178** (2002) 1053–1057
3. Jang, H., Lee, K., Kwon, O., Rhee, C., Shim, Y., Han, J.: Bronchioloalveolar carcinoma: focal area of ground-glass attenuation at thin-section CT as an early sign. Radiology **199** (1996) 485–488
4. Kuriyama, K., Seto, M., Kasugai, T., Higashiyama, M., Kido, S., Sawai, Y., Kodama, K., Kuroda, D.: Ground-glass opacity on thin-section CT: value in differentiating subtypes of adenocarcinoma of the lung. Am. J. Roentgenol. **173** (1999) 465–469
5. Watanabe, S., Watanabe, T., Arai, K., Kasai, T., Haratake, J., Urayama, H.: Results of wedge resection for focal bronchioloalveolar carcinoma showing pure ground-glass attenuation on computed tomography. Ann. Thorac. Surg. **73** (2002) 1071–1075
6. Heitmann, K., Kauczor, H.U., Mildenberger, P., Uthmann, T., Perl, J., Thelen, M.: Automatic detection of ground glass opacities on lung HRCT using multiple neural networks. European Radiology **7** (1997) 1463–1472
7. Tanino, M., Takizawa, H., Yamamoto, S., Matsumoto, T., Tateno, Y., Iinuma, T.: A detection method of ground glass opacities in chest x-ray CT images using automatic clustering techniques. In: SPIE. Volume 5032. (2003) 1728–1737
8. Zhang, L., Fang, M., Naidich, D., Novak, C.: Consistent interactive segmentation of pulmonary ground glass nodules identified in CT studies. In: SPIE. Volume 5370. (2004) 1709–1719
9. Chang, S., Emoto, H., Metaxas, D., Axel, L.: Pulmonary micronodule detection from 3d chest CT. In: MICCAI. (2004) 821–828
10. Freund, Y., Schapire, R.: Experiments with a new boosting algorithm. In: the 13^{th} International Conference on Machine Learning. (1996) 148–156
11. Huang, X., Qian, Z., Huang, R., Metaxas, D.: Deformable-model based textured object segmentation. In: EMMCVPR. (2005) 119–135
12. Lorenz, C., Carlsen, I.C., Buzug, T., Fassnacht, C., Weese, J.: Multi-scale line segmentation with automatic estimation of width, contrast and tangential direction in 2d and 3d medical images. In: CVRMed-MRCAS. (1997) 233–42
13. Ali, S., Silvey, S.: A general class of coefficients of divergence of one distribution from another. J. Roy. Stat. Soc. **28** (1966) 131–142
14. Chernoff, H.: Large-sample theory: Parametric case. Ann. Math. Stat. **27** (1956) 1–22
15. Zhou, J., Chang, S., Metaxas, D., Axel, L.: Vessel boundary extraction using ridge scan-conversion and the deformable model. In: ISBI. (2006) 189–192
16. Efron, B.: Estimating the error rate of a prediction rule: Improvement on cross-validation. Journal of the American Statistical Association **78** (1983) 316–331

Imaging of 3D Cardiac Electrical Activity: A Model-Based Recovery Framework

Linwei Wang[1], Heye Zhang[1], Pengcheng Shi[1,2], and Huafeng Liu[3]

[1] Department of Electrical and Computer Engineering,
Hong Kong University of Science and Technology, Hong Kong
maomwlw, eeship@ust.hk
[2] School of Biomedical Engineering, Southern Medical University, Guangzhou,China
[3] State Key Laboratory of Modern Optical Instrumentation, Zhejiang University,
Hangzhou, China

Abstract. We present a model-based framework for imaging 3D cardiac transmembrane potential (TMP) distributions from body surface potential (BSP) measurements. Based on physiologically motivated modeling of the spatiotemporal evolution of TMPs and their projection to body surface, the cardiac electrophysiology is modeled as a stochastic system with TMPs as the latent dynamics and BSPs as external measurements. Given the patient-specific data from BSP measurements and tomographic medical images, the inverse problem of electrocardiography (IECG) is solved via state estimation of the underlying system, using the unscented Kalman filtering (UKF) for data assimilation. By incorporating comprehensive *a priori* physiological information, the framework enables direct recovery of intracardiac electrophysiological events free from commonly used physical equivalent cardiac sources, and delivers accurate, robust, and fast converging results under different noise levels and types. Experiments concerning individual variances and pathologies are also conducted to verify its feasibility in patient-specific applications.

1 Introduction

Noninvasive imaging or recovery of cardiac electrical activity from BSPs would be highly beneficial both as a clinical aid and a mass screening tool. Conventional strategies in IECG problem usually follow a typical pattern. First, a simplified physical equivalent source is used as *implicit* constraints on the problem, which is either surface projection or regional accumulation of the realistic electrical activity. Then as the central focus of most recent IECG investigations, mathematically motivated regularization is imposed as additional *explicit* constraints. In these approaches, difficulties in interpreting the relationship of the solution and underlying physiopathology are involved, and the lack of physiologically meaningful constraints remains the bottleneck that hinders the direct recovery scheme free from physical equivalent sources.

More fundamentally in terms of physiology, it is recognized that time-varying 3D TMP distributions not only preserves characteristic electrophysiological

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 792–799, 2006.

dynamics but also directly relates to alterations in BSPs. However, although remarkable achievements have been made in the forward simulation of BSPs from TMPs [1,2], little success has been achieved in the inverse direction since the problem is highly under-determined. We believe that adequate constraints are necessary and the incorporation of known electrophysiological information is a desired paradigm. Although an *indirect* recovery method based on optimization of heart excitation model parameters, the important work [3] shares with our intuition that rich *a priori* information from physiological models would ensure the meaningfulness of the constraints and reduce the difficulty in explicitly applying multiple constraints.

In this paper, we present a framework for *directly* recovering cardiac electrical activity in terms of 3D myocardial TMPs. In the framework, the system of electrophysiology is carefully modeled, and 3D cardiac electrical activity is recovered via state estimation of the underlying system. The dynamics of the system is described by the spatiotemporal evolution of TMPs and its relations to measurements by the TMP-BSP projection model. A stochastic state space interpretation of the system is then obtained. With the availability of patient-specific data from BSP measurements and heart-torso geometry from tomographic medical images, the recovery of TMPs is achieved using the sequential estimation technique, the unscented Kalman filtering (UKF), which is particularly suitable for our system with highly nonlinearity and complex system and data uncertainties beyond typical Gaussian assumption. Volumetric TMP map sequence is then generated to provide intuitive visualization of the patient-dependent electrophysiological dynamics throughout the entire cardiac cycle. Experiments concerning individual variances and pathologies are conducted to verify its accuracy, robustness, and feasibility in practical applications.

2 Methodology

2.1 System Modeling of Cardiac electrophysiology

Motivated by using the *a priori* physiological knowledge to constrain the IECG problem, we carefully model the nonlinear dynamic system of cardiac electrophysiology. The model-based system approach, on one hand, impose comprehensive physiological spatiotemporal constraints on the inverse problem simultaneously; on the other, BSP measurements and medical images (MRI and CT) from specific subjects are utilized complementarily to provide patient-specific information in the inverse process. The modeling work consists of 3 major components:

1. The patient-specific, image-derived heart-torso geometry model.
2. The nonlinear dynamic model describing spatiotemporal evolution of TMPs.
3. The projection model directly relating TMPs to BSPs.

It is worthy to point out that at current stage, we concentrante on illustrating the basic idea and rationales of the innovate recovery framework. Thus

the modeling criteria emphasize *adequate* physiological constraints on the inverse recovery process, rather than *realistic* descriptions of electrophysiological phenomena. Without loss of generality, different models with varying complexity and differing specialization can be easily accommodated in the future work.

Image-derived heart-torso model. Patient-specific heart-torso model is constructed using geometry data extracted from tomographic medical images and proper numerical representations. Concerned with the intervention of myocardial anisotropy and inhomogeneity on intra-/extra-cardiac potential distribution [4], and its close relationship to pathology [2], the anisotropic and inhomogeneous heart model is used with mesh free representation, where the anisotropy is reflected via detailed fiber structure and the inhomogeneity via position-dependent electrical properties. In Contrast, since the inhomogeneity of the torso has little impact on BSP pattern [5], and experimental measurements [6] show that conductivity anisotropy is dominated by intracardiac tissue, the isotropic and homogeneous model based on surface (BEM) representation is a viable option.

Dynamic model for spatiotemporal evolution of 3D TMP. The dynamics of the electrophysiological system is described by the FithHugh-Nagumo-like action-diffusion equations since it maintains characteristic phenomenal properties of the dynamics of interests with computational feasibility.

With a Garlekin mesh free representation [7], the two-variable model can be written in compact form with matrices:

$$\frac{\partial U}{\partial t} = -M^{-1}KU + f(U,V) \tag{1}$$

$$\frac{\partial V}{\partial t} = b(U - dV) \tag{2}$$

$$f(U,V) = c_1 U(1-U)(U-a) - c_2 UV \tag{3}$$

where U is transmembrane potential field and V for inhibition. The mass matrix M is a function of material density and the stiff matrix K is a function of the passive electrical conductivity. $f(U,V)$ describes the excitation where a, b, c_1, c_2, d are parameters defining shapes of action potentials.

TMP-BSP projection model. The TMP-BSP projection model relates unobservable TMPs to noninvasive BSP measurements. In the quasi-static condition, the problem is viewed in a passive volume conductor with the source distributed only in the myocardium and the governing poisson equation with given boundary condition [1] is:

$$\nabla \cdot (\sigma \nabla \phi) = -\nabla \cdot (\sigma_h \nabla u) \tag{4}$$

where ϕ, u is for BSPs and TMPs, σ_h, σ for intracardiac and extracardiac conductivity. Applying boundary integral method to equation(4) we get an equation with both surface and volume integral:

$$c(\xi)\phi(\xi)+\int_{\Gamma_T}\phi(x)q^*(\xi,x)\,\mathrm{d}\Gamma_T-\int_{\Omega_H}(\nabla\cdot(\sigma_h\nabla u))\phi^*(\xi,x)\,\mathrm{d}\Omega_H=\int_{\Gamma_T}q(x)\phi^*(\xi,x)\,\mathrm{d}\Gamma_T \tag{5}$$

where ϕ^* and q^* are fundamental solutions with fixed forms. Γ_T, Γ_H stands for the body and heart surface, and Ω_H for myocardium volume.

Commonly, the volume integral in equation(5) is approximated with simplified distributed dipoles [1]. Alternatively we attempt an innovative mesh free approximation of the volume integral whereby 1), the use of simplified models is avoided; 2), through integral by part, the assumption which restricts TMPs within myocardium is automatically satisfied. To ensure unique solutions, constrains defining potential references are added [1] and the linear TMP-BSP relationship is established with Minimal Norm (MN) method:

$$\Phi=(H_a^TH_a)^{-1}H_a^TB_aU=CU \tag{6}$$

where H_a, B_a are augmented forms of matrices H, B, where H results from the boundary integral with the BEM and B from volume integral with the mesh free method.

2.2 3D TMPs Recovery Through State Estimation

Stochastic state space representation. By viewing TMPs as latent states and BSPs as external measurements, the electrophysiological system is represented in a state space representation. The dynamic model of 3D TMP evolution is transformed to the state equation and the TMP-BSP projection model to the measurement equation. Allowing for complex system and data uncertainties such as modeling errors, individual variances, system and measurement noises, which are too complicate to tackle with in the present study but too important to neglect, we rearrange equations(1,2,6) into stochastic state space representation with state vector $X(t)=(U(t)^TV(t)^T)^T$, measurement $Y(t)=\Phi(t)$, noise terms $\omega(t)$,$\nu(t)$ denoting uncertainties in dynamic and measurement models:

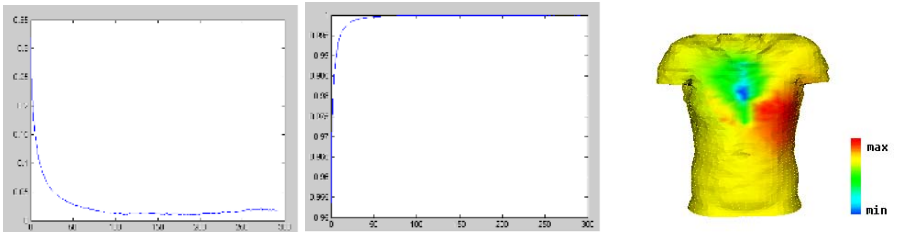
$$\frac{\partial X(t)}{\partial t}=\tilde{F}(X(t),\omega(t)) \tag{7}$$

$$Y(t)=\tilde{G}(X(t),\nu(t)) \tag{8}$$

where $\tilde{F}$ and $\tilde{G}$ represent functions from equations(1-3,6) after noises is incorporated[1]. Note that common additive Gaussian noise restrictions are not demanded, so complex uncertainties can be described in a more general way.

State estimation with the UKF. The system given by equations(7,8) has continuous dynamics but discrete measurements, thus further temporal discretization is demanded. A fourth-order Runge-Kutta solver to equation(7) is embedded in the inverse process to fulfil the discretization implicitly for the sake of reasonable accuracy and numeric stability.

[1] Detailed derivation omitted due to space limitation.



(a) (b) (c)

Fig. 1. Exemplary curves of RMSE (a) and CC (b) of recovery errors. (c), Exemplary BSP map at 140ms used as input for recovery framework.

Concerned with the nature of the system under study, which comprises highly nonlinear models without explicit discretized form and complicate uncertainties beyond Gaussian assumptions, we use the UKF as the minimum mean least square (MMLS) estimator for the state estimation [8]. In brief, TMPs are estimated in a recursive manner with each new measurement Y_k: 1), deterministically generate a set of points (sigma points) $\{\mathcal{X}_i^a, W_i\}_0^{2n}$ to approximate distribution of X_{k-1}^a based on previous estimates; 2), sigma points are transformed individually through state equation (7) and the prior distribution of the state is obtained; 3), the transformed sigma points are further transformed through measurement equation (8) and the distribution of the new measurement and its cross-covariance with the state are computed; 4), with the real measurement, the posterior distribution of the state is obtained through the linear update in a KF context.

In practical implementation, to reduce the sampling non-local effects [9] caused by the large dimension and strong nonlinearity of our model, and to improve efficiency, enhance numerical stability and guarantee positive semi-definite covariance matrices[10], square-rooted, scaled UKF is utilized.

3 3D Ventricular TMP Imaging: Results and Dicussion

Experiments are conducted on a specific heart-torso model constructed from geometry data provided by [11,12]. In the present study, to assess and exemplify the capabilities of the recovery framework, all results are from experiments on the phantoms, where longs run of the perfect models (7,8) are assumed to represent the ground truth. Observations are generated by adding noises of different types (Gaussian and Poisson) and different SNR levels (20dB and 30dB) to the true, downsampled BSPs. The generated BSP measurements are then feed back into the recovery framework, where the filtering technique is applied to estimate the latent dynamics of the electrophysiological system.

The resultant TMP map sequence represents complete cardiac electrical activity throughout the entire cardiac cycle. In early steps, individual variance or pathology is not considered. Later on, experiments concerned with patient-specific

Table 1. RMSE, CC and maximum point-point errors of the recovery errors after convergence in terms of Mean ± Deviance, with a initial value around 0.3 and 0.9 for RMSE and CC, and 0.5 for averaged scaled TMP value respectively

Error Metric	Gaussian(20dB)	Gaussian(30dB)	Poisson(20dB)	Poisson(30dB)
RMSE	0.0129±0.0022	0.0114±0.0018	0.0132±0.0033	0.0123±0.0020
CC	0.9999±e-005	0.9999±e-005	0.9999±e-005	0.9999±e-005
Max.error	0.0692±0.0050	0.0522±0.0033	0.0712±0.0058	0.0664±0.0052

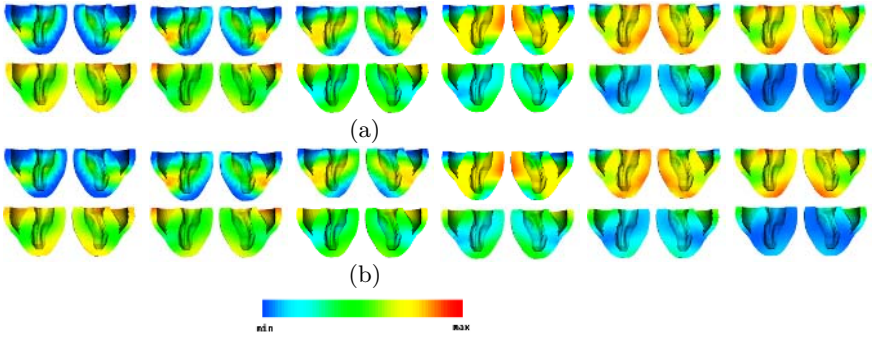


Fig. 2. Ventricular TMP map sequence of the ground truth (a) and the recovered results under 20dB Gaussian noise (b) at 12 time instances from 25ms to 300ms with 25ms interval between. All TMP values are scaled and all color scales go from minimum(blue) to maximum (red). In both sequence, the spread of the green front in the first 4 maps corresponds to the fast depolarisation (activation propagation), the spread of the light blue front in the last 4 maps represents the slower repolarisation (recovery sequence), and those between correspond to the plateau of TMPs.

case are also implemented to test the applicability of our framework. Assess of results is presented both qualitatively through TMP imaging and quantitatively in terms of global RMSE (relative mean square errors), CC (correlation coefficients) and local point-point errors.

3D Ventricle TMP imaging of normal heart. In these initial experiments, both phantoms and the recovery employ models with standard parameters (set according to [13] in present study). Earliest activation sites are set according to experimental observation [14]. An exemplary input BSP map is shown in Figure 1.*c*. Typical RMSE and CC curves are listed in Figure 1.*a* and 1.*b* to give a rough example of the accuracy, fast convergence and consistent stability of results. The quantitative comparisons in Table. 1 illustrate the robustness of our framework to consistently provide accurate results under different disturbances. The recovered TMP map sequence against the true TMP map under 20dB Gaussian noises are presented in Figure 2[2]. The recovered TMP map sequence is in good accordance

[2] The detailed interpretation is in the caption.

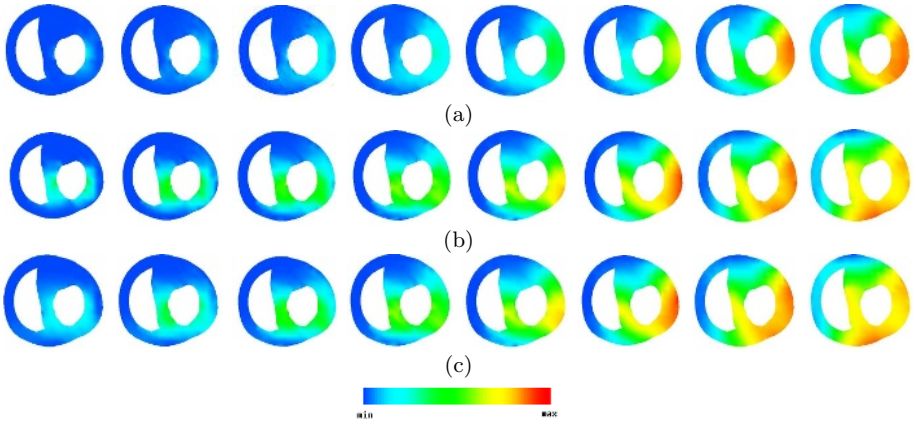


Fig. 3. TMP map sequence of cross section including the sites of ectopic focus at 8 time instances (Left to Right) from 15ms to 120ms with 15ms interval between: (a) the normal heart, (b) the ground truth of pathology, (c) the recovered TMPs for pathology

with the true sequence, both in spatial distributions and temporal evolutions, thus the capability of the 3D TMP imaging to provide accurate, comprehensive and intuitive visualization of cardiac electrical activity is verified.

TMP imaging for patient-specific cases. One desirable performance of the framework is its feasibility in dealing with pathologies and individual variances in patient-specific applications. The pathology of ectopic focus, caused by additional impulse besides those from sinoatrial nodes, is tested. The forward simulation is initiated by introducing two additional excitation points while the inverse recovery make no use of such *a prior* knowledge. The resultant RMSE and CC between recovered TMPs and pathologic ground truth is 0.0227 and 0.9997 while with normal ground truth is 0.1126 and 0.8024. Figure 3 lists the cross sectional map sequence of normal, pathologic and recovered TMP respectively. It verifies that though without *a prior* information, the recovery framework can not only timely and correctly reflect the cardiac abnormality, but also capture the site of ectopic focus. The future possibility of aiding clinical diagnosis can be expected.

Tests concerning individual variances are conducted with phantoms generated with slightly disturbed model parameters while the recovery implemented with standard ones. In an example considering conductivity variances[3], where longitudal component of σ_h is deviated from standard values by $\pm 10\%$ in phantom generation, the resultant RMSE and CC is 0.0122 ± 0.0028 and $0.9999 \pm$e-005 respectively. These first results are encouraging in the feasibility of the framework in patient-specific applications.

[3] Multitude experiments have been implemented investigating the influences of individual variances, however due to space limitation, only exemplary results are given in this paper.

Efforts focusing on improving the framework in patient-specific applications, primarily through automatically adjusting the generic model or its parameters to patient-specific data, and via data-fusion using electromechanical coupling, are under progress.

Acknowledgement. This work is supported in part by China National Basic Research Program (973-2003CB716100), Hong Kong Research Grants Council (CERG-HKUST6151/03E) and National Natural Science Foundation of China (60403040). For providing geometry data, we thank the Center for Integrated Biomedical Computing (CIBC) at the Univ. of Utah and the Bioengineering Institute of Auckland Univ. of Technology.

References

1. Aoki, M., Okamoto, Y., Musha, T., Harumi, K.: Three dimensional simulation of the ventricular depolarization and repolarization processes and body surface potentials: normal heart and bundle branch block. IEEE Trans.Biomed.Eng. **34** (1964) 454–461
2. Wei, D., Okazaki, O., Harumi, K., Harasawa, E., Hosaka, H.: Comparative simulation of excitation and body surface electrocardiogram with isotropic and anisotropic computer heart models. IEEE Trans.Biomed.Eng. **42** (1995) 343–357
3. He, B., Li, G., Zhang, X.: Noninvasive imaging of cardiac transmembrane potentials within three-dimdensional myocardium by means of a realistic geometry aniostropic heart model. IEEE Trans.Biomed.Eng. **50** (2003) 1190–1202
4. Roberts, D., Hersh, L., Scher, A.: Influence of cardiac fiber orientation on wavefront voltage, conduction velocity and tissue resistivity. Circ.Res. **44** (1979) 701–712
5. Rapport, B., Rudy, Y.: The inverse problem in electrocardiography: a model study of the effects of geometry and conductivity parameters on the reconstruction of epicardial pontetials. IEEE Trans.Biomed.Eng. **33** (1986) 667–675
6. Rush, S., Abildskov, J., R.McFee: Risistivity of body tissues at low frequencies. Circ.Res. **12** (1963) 40–50
7. Zhang, H., Shi, P.: A meshfree method for solving cardiac electrical propagation. In Ann.Int.Conf.IEEE EMBS (2005)
8. Julier, S., Uhlmann, J.: A new extension of the kalman filter to nonlinear systems. In Proc.of AeroSense:The 11th Int.Symp.on Aerospace/Defence Sensing,Simulation and Controls (1997)
9. Julier, S.: The scaled unscented transform. Int. J. Numer. Methods Eng. **47** (2000) 1445–1462
10. Merwe, R., Wan, E.: The square–root unscented kalman filter for state and parameter-estimation. ICASSP **6** (2001) 3461–3464
11. Nash, M.: Mechanics and material properties of the heart using an anatomically accurate mathematical model. University of Auckland (1998)
12. MacLeod, R., Johnson, C., Ershler, P.: Construction of an inhomogeneous model of the human torso for use in computational electrocardiography. In 13th Ann.Int.Conf. IEEE EMBS. (1991) 688–689
13. Rogers, J., McCulloch, A.: A collocation-galerkin finite element model of cardiac action potential propagation. IEEE Trans.Biomed.Eng. **41** (1994) 743–757
14. Durrur, D., Dam, R., Freud, G., Janse, M., Meijler, F., Arzbaecher, R.: Total excitation of the isolated human heart. Comp. Methods Appl. Mech. Eng. **41** (1970) 899–912

Segmentation of the Surfaces of the Retinal Layer from OCT Images

Mona Haeker[1,2], Michael Abràmoff[1,3], Randy Kardon[3], and Milan Sonka[1,3]

[1] Department of Electrical and Computer Engineering,
[2] Department of Biomedical Engineering,
[3] Department of Ophthalmology and Visual Sciences,
The University of Iowa, Iowa City, IA 52242, USA
{mona-haeker, michael-abramoff, randy-kardon, milan-sonka}@uiowa.edu

Abstract. We have developed a method for the automated segmentation of the internal limiting membrane and the pigment epithelium in 3-D OCT retinal images. Each surface was found as a minimum *s-t* cut from a geometric graph constructed from edge/regional information and *a priori*-determined surface constraints. Our approach was tested on 18 3-D data sets (9 from patients with normal optic discs and 9 from patients with papilledema) obtained using a Stratus OCT-3 scanner. Qualitative analysis of surface detection correctness indicates that our method consistently found the correct surfaces and outperformed the proprietary algorithm used in the Stratus OCT-3 scanner. For example, for the internal limiting membrane, 4% of the 2-D scans had minor failures with no major failures using our approach, but 19% of the 2-D scans using the Stratus OCT-3 scanner had minor or complete failures.

1 Introduction

Retinal thickness measurements obtained from optical coherence tomography (OCT) provides clinically useful information for the diagnosis of diseases of the optic nerve head, such as glaucoma [1]. However, to date, very little has been reported on the automated analysis of such images [2]. Furthermore, although "volumetric" data is often obtained by performing multiple circular scans (typically six), current methods rely on detecting the retinal boundaries in each 2-D cross-sectional image individually [1, 2]. To our knowledge, nothing has been reported about the automated detection of the *surfaces* of the retinal layers. This work presents a method for automatically detecting the surfaces of the retinal layer in 3-D OCT images: the internal limiting membrane and the pigment epithelium, thereby allowing quantitative measurements of retinal thickness. In particular, our method finds each surface by finding a minimum *s-t* cut in a 3-D geometric graph constructed from edge/regional information and *a priori*-determined surface constraints. The internal limiting membrane is found first and the pigment epithelium second, thereby allowing additional information to be used in the construction of the graph for detecting the pigment epithelial surface.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 800–807, 2006.

The process of transforming a multiple surface segmentation problem into that of finding a minimum s-t cut in a geometric graph has been previously reported by Li *et al.* [3]. It extends a previously reported method for detecting a single optimal surface by Wu and Chen [4] by adding additional edges to model interactions between the surfaces. While our method does not find the surfaces simultaneously as in [3], it imposes similar surface interaction constraints. One important advantage of using the minimum cost s-t cut surface detection method [3, 4] when compared to other previously-reported 3-D based surface segmentation methods [5–7] is that it guarantees to find the optimal solution with respect to the cost function.

2 Methods

2.1 The Sequential Multiple Surface Segmentation Problem

As OCT images were acquired as a sequence of circular scans (Fig. 1), the images naturally used a cylindrical coordinate system. Let $I(r, \theta, z)$ represent a 3-D OCT image, with corresponding image sizes R, T, and Z. R reflects the number of circular scans (typically 6), T reflects the number of angles used in each circular scan (typically 128), and Z reflects the number of points sampled at each (r, θ) point (typically 1024).

When using such a coordinate system, the surfaces of the retinal layer appear terrain-like. In particular, for each column of z-values corresponding to a specific (r, θ) pair, each surface intersects the column only once. Thus, each surface could be defined with a function $f(r, \theta)$, mapping (r, θ) pairs to z-values. Let $f_{ilm}(r, \theta)$ define the internal limiting membrane and $f_{rpe}(r, \theta)$ define the pigment epithelial surface of the retina. It then became the goal of our multiple surface segmentation method to determine $f_{ilm}(r, \theta)$ and $f_{rpe}(r, \theta)$.

Given a cost function $c(r, \theta, z)$ that specifies a surface unlikelihood measurement for each voxel, the cost of each surface was defined as the total cost of all voxels on the surface. An optimal surface was defined as the feasible surface for

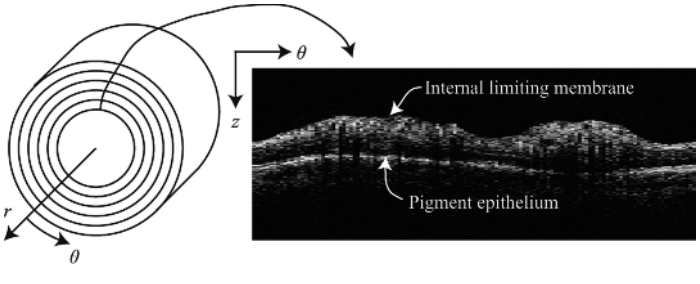


Fig. 1. Each 3-D OCT image is composed of a number of 2-D circular scans, as schematically shown on the left. One example circular scan is shown on the right. The internal limiting membrane is the "upper" surface of the retinal layer (smaller z-values) and the pigment epithelium is the "lower" surface of the retinal layer (larger z-values).

which the surface cost was minimal. With a careful transformation of the costs of the voxels as described in [3] or [4] into the weights of a directed 3-D graph, an optimal surface directly corresponded to the minimum-cost (nonempty) closed set in the graph. Note that a closed set was a subset of the graph such that no directed edges left the set. The minimum-cost closed set in the transformed graph could be determined by finding the minimum s-t cut of the graph.

Our method used a sequential approach in detecting both surfaces. In particular, after finding the first surface using the single-surface detection method, we used the information from the location of the first surface to aid in creating the geometric graph used for finding the second surface (e.g., to define a more appropriate cost function and to limit the search range of the second surface).

2.2 Surface Feasibility

It was the structure of each graph (vertices and edges) that guaranteed that a feasible surface would be found. Two major types of surface constraints were imposed: (1) smoothness constraints for each individual surface (i.e., how much the radial and theta values were allowed to vary from column to column) and (2) surface interaction constraints (i.e., the minimum and maximum allowed distances between the surfaces). In detecting the internal limiting membrane, only smoothness constraints were imposed since this surface was determined first. Once the internal limiting membrane was found, the interaction constraints could be added for detection of the pigment epithelial surface by limiting the search range.

More specifically, both surfaces were defined to be feasible if they satisfied the following properties:

- Smoothness constraint: The difference in the z-value between the points on the surface in adjacent columns did not exceed a constant value. In particular, $|f(r,\theta+1) - f(r,\theta)|$ was required to be less than or equal to $\Delta\theta$ for all (r,θ) and $|f(r+1,\theta) - f(r,\theta)|$ was required to be less than or equal to Δr for all (r,θ). Furthermore, since a larger physical distance was associated with each unit change of r, Δr was set to be greater than $\Delta\theta$. Note that because we were using cylindrical coordinates, columns with $\theta = T-1$ were defined as being adjacent to columns with $\theta = 0$ (circularity constraint).
- Interaction constraint: The distance between the two surfaces was at least δ^l voxels and at most δ^u voxels (i.e., $\delta^l \leq f_{rpe}(r,\theta) - f_{ilm}(r,\theta) \leq \delta^u$ for all (r,θ)).

2.3 Cost Function

Whereas the structure of the graph defined the set of feasible surfaces, it was the cost function itself that was integral in defining which surfaces were finally identified. Intuitively, the cost functions in this work were based on the following *a priori* observable properties about the retinal layer: a strong dark-to-light transition occurred at the internal limiting membrane, a dark-to-light transition

occurred at the pigment epithelial surface, very few voxels above the internal limiting membrane were bright, and the voxels immediately following (larger z-value) each surface had a higher intensity, on average, than the voxels preceding each surface. Reflecting these properties, the cost functions $c(r, \theta, z)$ used in this work had two major components: an edge-based term $e(r, \theta, z)$ and a region-based term $v(r, \theta, z)$:

$$c(r, \theta, z) = \alpha e(r, \theta, z) + \beta v(r, \theta, z), \tag{1}$$

where α and β were constant normalizing terms. Furthermore, as the two surfaces were expected to have somewhat different properties, a different cost function was used for each.

Image pre-processing. Like ultrasound images, OCT images contain speckle noise. Thus, as a pre-processing step, a speckle-reducing anisotropic diffusion (SRAD) method [8] was first applied to each circular scan image before computing each cost function. This was in an attempt to reduce the effects of the speckle noise, while maintaining edge information.

Edge information. A 3x3 Sobel kernel was convolved with each pre-processed image, resulting in an estimate the strength of the edge in the z-direction for each voxel (giving positive values for light-to-dark transitions and negative values for dark-to-light transitions). Because we expected a dark-to-light transition to occur at both surfaces, we maintained the signed edge values in the cost function, thus favoring dark-to-light transitions and discouraging light-to-dark transitions.

Regional information. The regional cost terms were added to help aid in the detection of the correct surface even when edge-information was lacking. For the internal limiting membrane, we used a normalized cumulative image as our regional cost term. This was based on the observation that very few bright voxels exist above (have a lower z-value than) the internal limiting membrane in each column. Incorporating a cumulative image into the cost also had the effect of discouraging the selection of voxels in the pigment epithelial surface, a necessity since we were finding the internal limiting membrane first.

For the pigment epithelial surface, we incorporated a region-based term into the cost function that attempted to minimize the regional variances above and below the surface. As an example, suppose there are only two relatively-homogeneous regions in a 3-D image separated by a surface S with average intensities a_1 (from voxel locations (x, y, z) inside S) and a_2 (from voxel locations (x, y, z) outside S). Then a regional cost function term as presented by Chan and Vese [9] can be defined as

$$\begin{aligned} F(S, a_1, a_2) = &\int_{inside(S)} |I(x, y, z) - a_1)|^2 \, dx \, dy \, dz \\ &+ \int_{outside(S)} |I(x, y, z) - a_2|^2 \, dx \, dy \, dz. \end{aligned} \tag{2}$$

In our work, we maintained the idea of minimizing the variance on either side of the surface, but we limited the computation to a region of limited size. This was because the assumption of having only two homogeneous regions in OCT images was only applicable to a limited area around each surface. Furthermore, because our cost function was voxel-based and not surface-based (the cost of a surface in this work was computed from a summation of voxel-based costs and thus could not include terms that depended on the overall surface properties), we added the variances of the voxels above and below each voxel for which we were computing the cost:

$$\widehat{v}(r,\theta,z) = \sum_{k=z-\gamma}^{z-1} (I(r,\theta,k) - \bar{I}_-)^2 + \sum_{k=z+1}^{z+\gamma} (I(r,\theta,k) - \bar{I}_+)^2, \tag{3}$$

where γ was a constant defining how far above and below the voxel we computed the variance, $\bar{I}_-$ was the average intensity of voxels with smaller z-values than the voxel at position (r,θ,z), and $\bar{I}_+$ was the average intensity of voxels with larger z-values than the voxel at position (r,θ,z). However, since Eq. (3) would also favor voxels that were not of interest (e.g., background voxels), the actual regional cost term $v(r,\theta,z)$ would use the cost defined by Eq. (3) only if $\bar{I}_+$ were greater than $\bar{I}_-$:

$$v(r,\theta,z) = \begin{cases} \widehat{v}(r,\theta,z) & \text{if} \quad \bar{I}_+ + \varepsilon > \bar{I}_- \\ c & \text{otherwise,} \end{cases} \tag{4}$$

where c was a relatively large constant, thereby giving a large cost to potential surface voxels that do not separate a high intensity region "below" (larger z-values) from a low intensity region "above" (smaller z-values) the voxel.

3 Experimental Methods

The method was tested on 18 3-D OCT data sets obtained from a Stratus OCT-3 scanner: 9 were from patients with normal optic discs and 9 were from patients with papilledema (optic nerve swelling). The dimensions of each data set were 6x128x1024 ($R = 6$, $T = 128$, $Z = 1024$). The axial scan length was 2 mm, resulting in a voxel size of approximately 2 μm in the z-direction. The size of the voxels in the r and θ varied due to the radial nature of the scans. The following surface constraints were used:

- $\Delta\theta = 10$ pixels for both surfaces,
- $\Delta r = 150$ pixels for the internal limiting membrane, $\Delta r = 100$ pixels for the pigment epithelial surface,
- $\delta^l = 60$ pixels (minimum distance between surfaces), and
- $\delta^u = 600$ pixels (maximum distance between surfaces).

Each resulting surface was plotted on the six corresponding circular scans and its correctness assessed by an expert observer. In addition, the internal limiting

membrane resulting from our method was compared to that of the method used by Stratus OCT-3 system (Stratus OCT-3 results for the pigment epithelium were not available).

4 Results

Expert analysis of all the data sets indicated that our method found the correct surfaces. Fig. 2 displays an example of typical segmentation results. The corresponding retinal thickness plots are shown in Fig. 3.

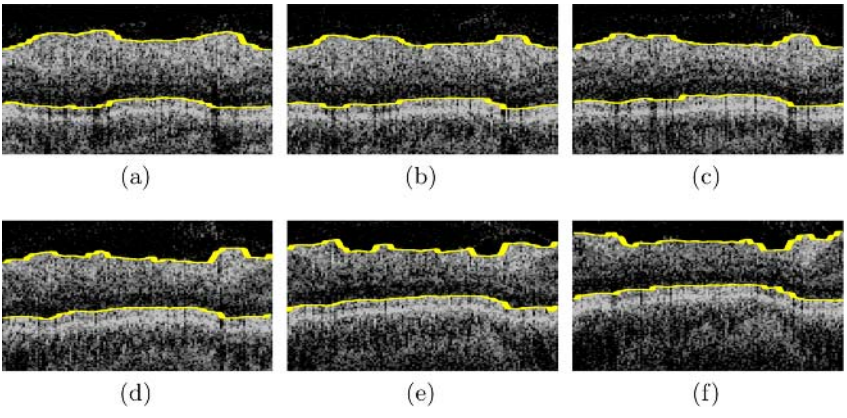


Fig. 2. Typical results for one 3-D data set. The results are shown in increasing circular scan order such that the results for the inner circular scan are shown in (a), while the results for the outer circular scan are shown in (f). Note that for easier visualization of the borders, the top and bottom of the images have been cropped.

Table 1 shows a comparison of the "failures" of our method versus that of the Stratus OCT-3 for the internal limiting membrane. The counts given in the table refer to the number of 2-D circular scans on which some type of failure was found out of all of the circular scans for each scan group (54 normal scans, 54 papilledema scans, 108 total scans). A circular scan was considered to have a "minor failure" if the visible deviation of the result from the correct surface was small enough not to require correction in a clinical setting. A scan was considered to have a "complete failure" if the result had a large deviation that would require correction in order for the segmentation to be clinically useful. The Stratus OCT-3 system had two such failures, as shown in Fig. 4 (our results are also shown in the figure for comparison). Minor failures (4/108) of our method occurred in cases in which the image data did not satisfy our circularity assumption. Note that we consider our result in Fig. 4 (c) our worst "minor failure." Each example is from an inner circular scan from a patient with papilledema.

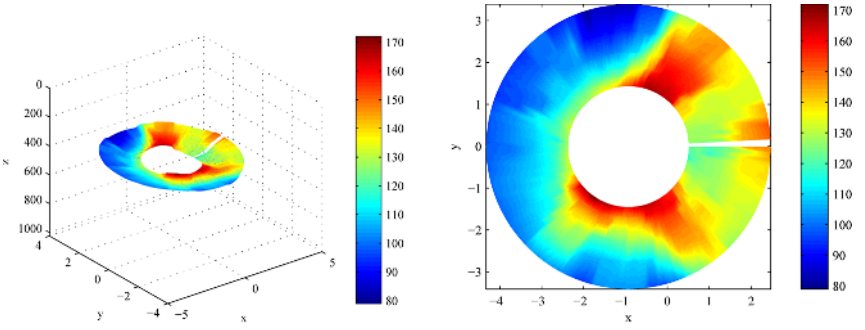


Fig. 3. Color-coded retinal thickness plots corresponding to results shown in Fig. 2. (a) 3-D view of the internal limiting membrane with colors reflecting retinal thickness (in pixels). (b) 2-D view of retinal thickness values (in pixels).

Table 1. Count of circular scan failures in our method and the Stratus OCT-3 method for the internal limiting membrane

Scan group	Method	Minor failures	Complete failures	Total failures
Normal scans	3-D Graph Search	0 (0.0%)	0 (0.0%)	0 (0.0%)
Normal scans	Stratus OCT-3	4 (7.4%)	0 (0.0%)	4 (7.4%)
Papilledema scans	3-D Graph Search	4 (7.4%)	0 (0.0%)	4 (7.4%)
Papilledema scans	Stratus OCT-3	15 (27.8%)	2 (3.7%)	17 (31.5%)
All scans	3-D Graph Search	4 (3.7%)	0 (0.0%)	4 (3.7%)
All scans	Stratus OCT-3	19 (17.6%)	2 (1.8%)	21 (19.4%)

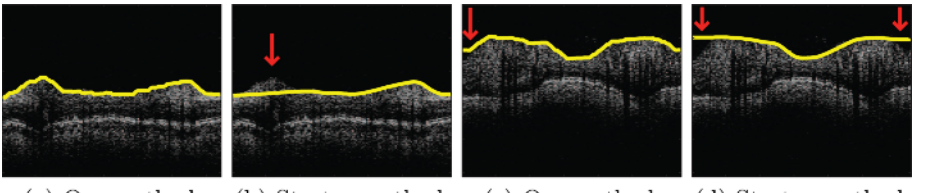


Fig. 4. Examples of "complete failures" (arrows) for the Stratus OCT-3 method and our corresponding results for the internal limiting membrane. (a)–(b) Our method versus the Stratus OCT-3 method for an inner circular scan on a patient with papilledema, (b) is considered a "complete failure." (c)–(d) Our method versus the Stratus OCT-3 method for an inner circular scan on a different patient with papilledema, (c) is considered a "minor failure" and (d) is considered a "complete failure."

5 Discussion and Conclusion

The presented qualitative analysis of the results from all 18 3-D OCT data sets (108 total circular scans) indicates that our method identified the correct

border for the internal limiting membrane in 96% of the circular scans (93% in papilledema scans). This offers a substantial improvement over the Stratus OCT-3 system which only found the correct border 81% of the time (69% in papilledema scans). The 4/108 minor failures of our method (e.g., see Fig. 4 (c)) can be attributed to the invalidity of the circularity constraint in these cases.

We have presented a fully automated 3-D method for detecting the internal limiting membrane and pigment epithelial surface of the retina from OCT images that we hope to evaluate on clinical datasets.

References

1. Jaffe, G.J., Caprioli, J.: Optical coherence tomography to detect and manage retinal disease and glaucoma. Am. J. Ophthalmol. **137**(1) (2004) 156–169
2. Koozekanani, D., Boyer, K., Roberts, C.: Retinal thickness measurements from optical coherence tomography using a Markov boundary model. IEEE Trans. Med. Imag. **20**(9) (2001) 900–916
3. Li, K., Wu, X., Chen, D.Z., Sonka, M.: Optimal surface segmentation in volumetric images – a graph-theoretic approach. IEEE Trans. Pattern Anal. Machine Intell. **28**(1) (2006) 119–134
4. Wu, X., Chen, D.Z.: Optimal net surface problems wih applications. In: Proc. of the 29th International Collouium on Automata, Languages, and Programming (ICALP), LNCS 2380, Springer-Verlag (2002) 1029–1042
5. Thedens, D.R., Skorton, D.J., Fleagle, S.R.: Methods of graph searching for border detection in image sequences with applications to cardiac magnetic resonance imaging. IEEE Trans. Med. Imag. **14**(1) (1995) 42–55
6. Suri, J.S., Liu, K., Singh, S., Laxinarayan, S.N., Zeng, X., Reden, L.: Shape recovery algorithms using level sets in 2-D/3-D medical imagery: A state-of-the-art review. IEEE Trans. Inform. Technol. Biomed. **6**(1) (2002) 8–28
7. McInerney, T., Terzoppoulos, D.: Deformable models in medical image analysis: A survey. Medical Image Analysis **1**(2) (1996) 91–108
8. Yu, Y., Acton, S.T.: Speckle reducing anisotropic diffusion. IEEE Trans. Image Processing **11**(11) (2002) 1260–1270
9. Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Trans. Image Processing **10**(2) (2001) 266–277

Spinal Crawlers: Deformable Organisms for Spinal Cord Segmentation and Analysis

Chris McIntosh and Ghassan Hamarneh

Medical Image Analysis Lab, School of Computing Science Simon Fraser University, BC, Canada
{cmcintos, hamarneh}@cs.sfu.ca

Abstract. Spinal cord analysis is an important problem relating to the study of various neurological diseases. We present a novel approach to spinal cord segmentation in magnetic resonance images. Our method uses 3D "deformable organisms" (DefOrg) an artificial life framework for medical image analysis that complements classical deformable models (snakes and deformable meshes) with high-level, anatomically-driven control mechanisms. The DefOrg framework allows us to model the organism's body as a growing generalized tubular spring-mass system with an adaptive and predominantly elliptical cross section, and to equip them with spinal cord specific sensory modules, behavioral routines and decision making strategies. The result is a new breed of robust DefOrgs, "spinal crawlers", that crawl along spinal cords in 3D images, accurately segmenting boundaries, and providing sophisticated, clinically-relevant structural analysis. We validate our method through the segmentation of spinal cords in clinical data and provide comparisons to other segmentation techniques.

1 Introduction

The spinal cord is a crucial part of the nervous system that resides within the vertebral canal of the spinal column and acts as a relay to convey information between the brain and the rest of the body. However, there are numerous clinical problems related to the spinal cord, including multiple sclerosis (MS), meningitis, neural tube defects, syringomyelia, transverse myelitis, and spinal cord injury (SCI). Of specific interest, MS affects more than a quarter million people in the US, and is suspected of shrinking the cord. A necessary prior for statistically analyzing the spinal cord is segmenting it. For example, [1] performs texture analysis on manually segmented spinal cords to study the effects of MS.

However, spinal cord segmentation is a difficult problem. Though the cord is approximately cylindrical in shape, its diameter varies at different vertebral levels. Furthermore, progressing from superior to inferior the information transfer requirement decreases resulting in decreased white matter and, therefore, decreased contrast. This change in contrast proves difficult for nonadaptive segmentation methods. Consequently, the majority of previous work focuses on segmenting the canal and the vertebra [2, 3, 4] (which is a much simpler problem given the high contrast of bone in CT).

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 808–815, 2006.

There are, however, a few existing works on spinal cord segmentation. In [5] Schmit and Cole investigate if an observable change in the spinal cord occurs after SCI. Segmentation is performed using seeded region growing, concluding that spinal narrowing in magnetic resonance images (MRI) is indeed observable. However, no validation of the segmentation results is presented, and the assumption that extraspinal structures do not touch the spinal cord is made, which often does not hold (Fig 3). In [6] a deformable model approach is used for spinal cord segmentation. However, the method requires initializations along the entire cord, is only validated on a single phantom data set, and takes approximately 10 hours for one segmentation. In [7] Aylward *et al.* use a hessian-based approach for strictly tubular (vessel) structures and, thusly, do not address the variability in the major and minor axes of the cord. Furthermore, no quantitative validation of the proposed technique on spinal cord data is provided. As we motivate in section 2.1 this is not an optimal approach as the spinal cord can vary from tubular to elliptical in shape, and local extraspinal structures distort derivatives of Gaussians along the minor-axis of the cross-sections.

In previous work we demonstrated the merit of locally optimal filtering in vessel segmentation and analysis [8]. In this work, we address the clinical need for a fast semi-automatic spinal cord segmentation and analysis technique by extending and adapting our previous work for this new problem. We demonstrate the use of hessian-based filters for spinal cord segmentation that intuitively encode the geometrical properties of the cord, along with an elliptical mesh that explicitly encodes the major and minor axes.

The remainder of the paper is structured as follows. We begin with an overview of our earlier work on DefOrgs for the segmentation and analysis of tubular structures. We then describe our method of spinal cord segmentation using optimal hessian scales and our extension to adaptive, elliptical, cross-sectional filtering (section 2). In section 3 we present results comparing our method to other to state-of-the art filtering and segmentation techniques. Finally, we summarize and draw conclusions in section 4.

1.1 Tubular Crawlers

In [8] we presented a novel approach to the segmentation and analysis of vasculature from volumetric medical image data. Our method extended the original DefOrgs introduced by Hamarneh *et al.* in [9] to 3D, modelled their bodies as tubular spring-mass systems and equipped them with a new repertoire of sensory modules, behavioral routines, and decision making strategies specifically designed for vascular segmentation. The result was a new breed of robust DefOrgs, "vessel crawlers", that crawl along vasculature in 3D images, accurately segmenting vessel boundaries, detecting and exploring bifurcations, and providing sophisticated, clinically-relevant structural analysis. However, our methods were restricted to high-contrast tubular structures on dark backgrounds (vessels) and are, thus, ill-suited for segmenting the spinal cord; a low-contrast elliptical structure.

2 Methods

In this section we provide an overview of the spinal crawler and its use in the segmentation and analysis of spinal cords in 3D MRI. We begin with an overview of the geometrical properties of the spinal cord, then describe the spinal crawler, followed by details of our implementation of the various layers of the DefOrgs framework.

2.1 Geometrical Properties of Spinal Cords

As previously mentioned, the cross-section of the spinal cord is tubular in some regions and elliptical in others (Fig 1). This geometrical property motivates the use of an adaptive filtering procedure that takes advantage of the local geometry. Here we describe such a procedure.

It is well known that the eigenvalues of the Hessian $H(I*G_\sigma)(x)$ computed at scale σ describe the principle curvature at location x [10]. Given an ordering of eigenvalues $|\lambda_1| \leq |\lambda_2| \leq |\lambda_3|$ the ratios between the eigenvalues yield measurements of the underlying geometrical structure. This property motivates Frangi *et al.*'s vesselness measure at a single scale, under which R_A differentiates between plates and lines, R_B measures deviation from blob like structures, and S emphasizes areas of high contrast (for details see [11]). The final filtering result at each voxel is the maximum of a search across scales with a fixed value for c, which is ill-suited to the spinal cord as it changes shape and dims along its medial axis.

In order to adapt Hessian based approaches to spinal cord segmentation, we locally adapt the gradient calculation method and filtering equations. We define an anisotropic Hessian $H(I * G_{\sigma_x,\sigma_y,\sigma_z})(x)$ computed at scales $\sigma_x, \sigma_y, \sigma_z$ where $G_{\sigma_x,\sigma_y,\sigma_z}$ is an anisotropic Gaussian with σ_x, σ_y equal to the major and minor axes of the elliptical structure and σ_z the average of the two; a somewhat arbitrary choice indicating how far the structure is expected to remain straight. However, a direct requirement of this type of filtering is that the major and minor axes of the structure be known. Furthermore, the eigenvalues are based on the difference taken inside and outside the structure and, as such, are influenced by the bright extraspinal structures that lie (anterior to posterior) around the spine (Fig 1). Consequently, we define a *'spinalness'* filter as:

$$S_{\sigma_{x,y,z}}(s) = \left\{ \begin{array}{c} 0 \; if \quad \lambda_2 > 0 \quad or \quad \lambda_3 > 0 \\ \left(e^{\left(-\frac{(R_A - 0.5)^2}{2\alpha^2} \right)} \right) e^{\left(-\frac{R_B^2}{2\beta^2} \right)} \left(1 - e^{-\frac{S^2}{2c^2}} \right) \quad o/w \end{array} \right\} \tag{1}$$

in which R_A, the ratio of the largerest two eigenvalues, is allowed to deviate from its expected value of *0.5*. In our framework we derive locally optimal values for σ_x, σ_y and c (section 2.2), thereby greatly enhancing the robustness of the method to image noise by adapting to the local cord shape and intensity level.

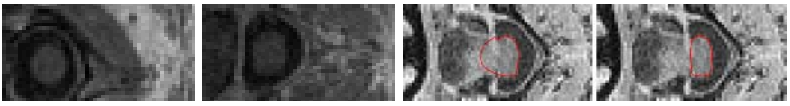


Fig. 1. Left pair: Axial slices of a spinal cord showing circular and elliptical cross-sections. Right pair: Simple user interaction on a cross-sectional slice can be used to repair false boundary identification. Right pair also shows the extraspinal structures, with gray level intensity higher than that of the spinal cord, that are often encountered in the minor-axis derivatives (horizontal-direction).

2.2 Spinal Crawlers

Our spinal crawler DefOrg was built under a multilevel Artificial Life modelling approach consisting of four primary levels: cognitive, behavioral, physical, and geometrical (Fig 2) [9]. Specifically, the cognitive layer makes decisions based on our model's current state, anatomical knowledge, and its surrounding environment (the image volume including the spinal cord). Decisions could be made to sense information, to deform based on sensory data, or to repair an erroneous layer. All of these actions are described under the behavioral layer of the organism, and they rely upon both the physical and geometrical layers for implementation. For example, the act of progressing down the spine is described by the *'growing'* behavioral method. The cognitive center gathers sensory input using the *'spine sense'* sensory module, decides the correct location via the *'grow'* decision module, elicits the act of *'growing'* , and then conforms to the spinal cord by *'fitting'*. In turn, each of these methods relies upon the physical and geometrical layers to carry out tasks, such as maintaining model stability through the application of stabilization springs (Fig 2 middle three). Consequently, we have a framework with many independent layers of abstraction, each built upon the implementation of independent models and processes. What follows is an explanation of each layer.

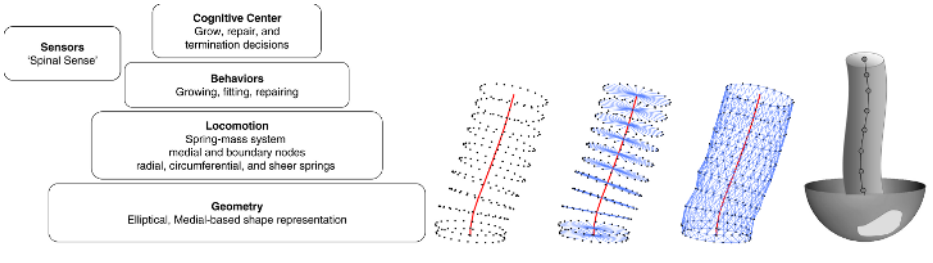


Fig. 2. Left: The layered architecture of DefOrgs showing all of the modules used by our spinal crawler. Middle three: Topology of the spinal crawler showing: Masses (left), radial springs (middle), and stability springs (right) across sequential layers of the organism. Right: A spinal crawler (gray) utilizing a hemispherical "off-board" (detached from body) sensor showing an idealized output (light gray).

Geometrical. The elliptical geometric module is parameterized by both section length (distance between neighboring medial masses) and circumferential resolution (proportional to the number of circumferential boundary masses) (Fig 2). Each layer supports major and minor axes lengths, and thus, is able to conform locally to the spinal cord.

Physical. We use a 3D Newtonian physically-based deformation module [8]. External forces are provided by the volumetric image gradient and a drag force, while internal forces are supplied through Hooke's law and dampening spring forces. This physical layer enables intuitive user-interaction, under which the user can interact with individual nodes to deform the entire system in a natural manner. This enables a quick and easy method of repairing erroneous segmentations in areas of non-existent boundary information (Fig 1)[12].

Sensory. The primary sensory module is controlled by the spinal crawler's decisions, wherein optimal parameters are derived through basic sensory modules (local image intensity, standard deviation, etc.). As noted in section 2.1 the Hessian and its associated filters must be computed at locally optimal scales. Our approach estimates the scales from the major and minor axes of the leading (front-most) layer of the spinal crawler after it has deformed to the local gradient information.

Specifically, our **'Spine Sense'** module captures volume intensity information on the surface of a hemisphere centered around the spinal crawler and normal to its front-most layer (Fig 2). A locally optimal connected-components filter is used to ensure only relevant information to the crawler is present on the surface. The filter returns all connected voxels with intensities and *'spinalness'* measurements that lie within $\pm$ multiples of standard deviations of maximum-likelihood-estimates of Gaussian distributions' means sampled from within the local segmented region of the spinal cord. Intensities from the filter now lie on the surface of the hemisphere, and the next location to grow to is taken as the centermost centroid of all circular regions on the surface. As shown in figure 3, this works quite well at sequentially detecting the local medial axis of the spinal cord.

Cognitive. The spinal crawler can make three key *decisions*. Namely, where to grow, when to repair, and when to stop. Each decision can be based on sensory input, anatomical knowledge, or user interaction. The user is able to override any of these key decision functions at any time during the organisms life cycle, and hence can illicit intuitive real-time control over the segmentation process. This presents a strong advantage over those methods that force the user to wait for a potentially incorrect output before re-running the segmentation process with different parameters (region growers, level sets, neural networks, etc.).

The **'Grow'** decision dynamically decides the next grow location via the *spinal sense* module, and activates the *growing* behavior with appropriate parameters (major and minor axis lengths, location for next layer, etc.). **'Repair'** decides to repair a given layer if its eccentricity exceeds the average eccentricity over

the past few layers. While **'Terminate'** decides when to terminate execution, which can be at the request of the user, when reaching a target point, or when a provided threshold of spinalness is not met.

Behavioral. Each of the spinal crawlers key decisions results in the execution of the appropriate *behavior* using the concluded locally optimal parameters, scales, estimated spinalness mean and variance, etc. **'Growing'** creates new layers and subsequently connects them to the current front-most layer (Fig 2). Once connected the model can be fit to the image data. **'Repairing'** takes advantage of the spinal cord's symmetric reflection about its major axis. The section of a layer exhibiting the highest eccentricity is replaced by a reflection of the opposite side. **'Fitting'** is accomplished using 3D image gradient driven deformations simulated by the physics layer. Connections to the previous layer provide smoothness, while stiffer circumferential springs provide local stability to noise, and flexible radial springs allow deformation to the spinal boundary.

3 Results

We present quantitative validation of our method on four $256 \times 256 \times 60$ MRIs, each of which has been manually segmented at approximately 60 minutes a volume [1], and has as a voxel size of $0.98 \times 0.98 \times 1mm^3$. As a point of comparison we also segment the data using connected components and ITK-SNAP's level-set method [13]. The connected components proved ill-suited to the task, even with heavy manual intervention, as the intensity levels of nearby structures were to similar to that of the cord (Fig 1, 3 far right). However, the level-set from SNAP was partially successful with high-degrees of manual intervention. Specifically, we had to partition the volume into pieces and segment each individual piece with slightly different parameters and seed point placements along most of the spine. Furthermore, in regions where boundaries are not clear, we had to re-process the image to create an improved edge image. Essentially, segmentation with SNAP required manually performing the automatic features of our framework. Overall, the whole process took on average 30 minutes for a single image. We expect this time will be even higher for a clinical expert with minimal knowledge of image segmentation algorithms and their parameters. Finally, the output from these techniques is simply a binary mask that itself cannot be intuitively interacted with nor easily analyzed (for example to obtain cross-sectional slices normal to the medial axis of the extracted spinal cord, skeletonization and pruning must be performed).

In contrast, segmentation with our method required on average 10 minutes per image with the user only providing two seed points. As can be shown in table 3 our method outperforms the SNAP method. Furthermore, as noted in section 2.2, our final results can be easily improved as shown in figure 1. In addition to a binary mask, the final segmentation without post-processing presents the

[1] Reported times performed by a medical image analysis software literate person.

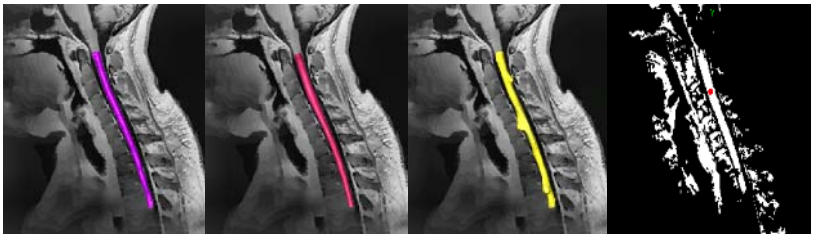


Fig. 3. Sample 3D segmentation results (left to right): manual, spinal crawler, SNAP level set, and a region grower with seed point shown in red that has leaked into nearby extraspinal structures

user with the medial axis, and cross-sectional slices of the spine taken along its normal[2]. Typically, 10 to 15 cross-sections of approximately 200 taken along the spinal cord were found in noticeable error, and can easily be corrected within a few minutes.

Table 1. Error comparison between the 2-seed-point, automatic spinal crawler, and the heavily adjusted/tuned SNAP. Errors are measured as minimal inter-surface distance computed against manual segmentations. Statistics are calculated across all inter-surface pairs and the 4 data sets. For example, the mean Hausdorff distance (maximum of inter-surface pairwise distances) across all 4 data sets for the spine crawler was 5.39.

Error	**Mean**	**Median**	**Min**	**Max**	**Std**
Spine Crawler Mean distance	0.5662	0.455	0.26	1.095	0.3694
SNAP Mean distance	0.812	0.734	0.415	1.39	0.4252
Spine Crawler Max error	5.39	5.485	4.1	6.5	1.07
SNAP Max error	5.71	4.83	3.58	9.6	2.73

4 Conclusion

We set out to develop a semi-automatic tool for spinal cord segmentation suitable for clinical studies of spinal cord-related diseases [1, 5]. To this end, we have presented an elliptical DefOrg that incorporates the cord's geometrical properties to delineate its medial axis and boundary. Our method has proven faster and more accurate then other segmentation techniques (connected components, level sets). Furthermore, through its intuitive and interactive control structures, real-time analysis, and post-segmentation interactions, it keeps the doctor *'in the loop'*; in contrast to numerous black box approaches that provide no interaction, no recourse in case of error, and whose inner-workings are nonintuitive.

[2] See http://mial.fas.sfu.ca/researchProject.php?s=157 for a video and larger images.

Acknowledgements

We would like to thank the MS/MRI Research Group at the University of British Columbia and Berlex for providing the spinal cord MR images, Lisa Tang for running tests on various data sets, and funding from the Natural Sciences and Engineering Research Council of Canada.

References

1. Mathias, J., Tofts, P., Losseff, N.: Texture analysis of spinal cord pathology in multiple sclerosis. Magnetic Resonance in Medicine **42** (1999) 929:935
2. D'Haese, P., Niermann, K., Cmelak, A., Duay, V., Li, R., Dawant, B.: Atlas-based spine segmentation for radiotherapy planning. American Society of Therapeutic Radiology and Oncology (2004)
3. Mancas, M., Gosselin, B., Macq, B.: Segmentation using a region-growing thresholding. Proceedings of the SPIE **5672** (2005) 388–398
4. Karangelis, G., Zimeras, S.: An accurate 3d segmentation method of the spinal canal applied to CT data. In: Bildverarbeitung fur die Medizin. (2002) 370–373
5. Schmit, B.D., Cole, M.K.: Quantification of morphological changes in the spinal cord in chronic human spinal cord injury using magnetic resonance imaging. IEEE EMBS **26** (2004) 4425 – 4428
6. Coulon, O., S.J.Hickman, Parker, G., Barker, G., Miller, D., Arridge1, S.: Quantification of spinal cord atrophy from magnetic resonance images via a b-spline active surface model. Magnetic Resonance in Medicine **47** (2002) 1176–1185
7. Aylward, S., Bullit, E., Pizer, S., Eberly, D.: Intensity ridge and widths for tubular object segmentation and description. IEEE/SIAM workshop Mathematical Methods Biomedical Image Analysis (1996) 131–138
8. McIntosh, C., Hamarnerh, G.: Vessel crawlers: 3d physically-based deformable organisms for vasulature segmentation and analysis. IEEE Conference on Computer Vision and Pattern Recognition **1** (2006) 1084–1091
9. Hamarneh, G., McInerney, T., Terzopoulos, D.: Deformable organisms for automatic medical image analysis. In: MICCAI. (2001) 66–76
10. Lindeberg, T.: Scale-Space Theory in Computer Vision. Kluwer Academic Publishers, Norwell, MA, USA (1994)
11. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale vessel enhancement filtering. Lecture Notes in Computer Science **1496** (1998) 130–137
12. Hamarneh, G., McIntosh, C.: Physics-based deformable organisms for medical image analysis. SPIE Medical Imaging **5747** (2005) 326–335
13. Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., Gee, J.C., Gerig, G.: User-guided 3d active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. Neuroimage (2006)

Markerless Endoscopic Registration and Referencing

Christian Wengert[1], Philippe C. Cattin[1], John M. Duff[2], Charles Baur[3], and Gábor Székely[1]

[1] Computer Vision Laboratory, ETH Zurich, 8092 Zurich, Switzerland
{wengert, cattin, szekely}@vision.ee.ethz.ch
[2] University Hospital of Lausanne, 1000 Lausanne, Switzerland
[3] Virtual Reality and Active Interfaces, EPFL Lausanne, 1015 Lausanne, Switzerland

Abstract. Accurate patient registration and referencing is a key element in navigated surgery. Unfortunately all existing methods are either invasive or very time consuming. We propose a fully non-invasive optical approach using a tracked monocular endoscope to reconstruct the surgical scene in 3D using photogrammetric methods. The 3D reconstruction can then be used for matching the pre-operative data to the intra-operative scene. In order to cope with the near real-time requirements for referencing, we use a novel, efficient 3D point management method during 3D model reconstruction.

The presented prototype system provides a reconstruction accuracy of 0.1 mm and a tracking accuracy of 0.5 mm on phantom data. The ability to cope with real data is demonstrated by cadaver experiments.

1 Introduction

Over the past decade computer aided, navigated surgery has evolved from early laboratory experiments to an indispensable tool for many interventions in spine surgery [1]. Accurate patient registration with the pre-operative imaging modalities or referencing of the target anatomical structure is an essential component of these procedures. Many of today's approaches are invasive and use either mechanical fixation devices, fiducial markers, ionizing radiation or require tedious manual registration with a pointing device. While the potential of ultrasound imaging for non-invasive registration and referencing is currently explored [2] its usability is constrained by the presence of air-filled cavities in the body.

In this paper we present a novel approach that uses a tracked monocular endoscope to register and reference vertebrae during spine surgery. By tracking natural landmarks over multiple views, a 3D reconstruction of the surgical scene can be computed using photogrammetric methods. The proposed algorithm consecutively refines the 3D model with each additional view. This reconstruction provides quantitative metric data about the target anatomy which can be used for 3D-3D registration of the anatomy to the pre-operative data and for further referencing. Our aim is to create an intra-operative support environment, which

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 816–823, 2006.

relies as much as possible on the tools and instrumentation used anyway during surgery and does not increase the invasiveness of the intervention just for navigation purposes.

The presented method is general and can be used for any rigid anatomical structure with sufficiently textured surfaces. In this paper, however, we concentrate on C2 vertebra referencing during C1/C2 transarticular screw placement, as this open intervention allows us to thoroughly test and validate our methods before applying them in minimal invasive spine surgery.

Various approaches have been proposed to provide the surgeon with more information during endoscopic interventions. Most of them are tailored to specific procedures and are of limited general applicability. In particular the following application fields can be identified: (a) registration and referencing, (b) navigation aids, and (c) augmented reality.

A fiducial marker based system to register freehand 2D endoscopic images to pre-operative 3D CT models of the brain has been presented in [3], and in [4] Thoranaghatte proposed a markerless referencing method for the spine. The use of markers, however, often forces the surgeon to increase the invasiveness of the intervention just to be able to perform referencing during the surgery. In [5], a hybrid tracking system is proposed using a magnetic tracking sensor and image registration between real and virtual endoscopic images to compute the pose and position of the camera in the CT coordinate frame.

A non-tracked calibrated endoscope for 3D reconstruction and motion estimation from endo-nasal images is used in [6] for registering the CT to the endoscopic video. Another navigation aid using photogrammetry during endoscopic surgery has been studied in [7]. They use the structural information to prevent the endoscope image from flipping upside-down while rotating the camera. In [8] the pose of a specially marked tool inside the surgical scene has been determined from monocular laparoscopic images and used to create 3D renderings from different views for the surgeon.

Augmented reality systems are not necessarily aiming at quantitative measurements but rather want to improve the visual perception for the surgeon by extending the image with additional information. In [9,10,11,12,13] externally tracked cameras are used to augment the surgeon's view by fusing pre-operative data with the actual endoscopic view. In contrast, [14] uses a stereo endoscope instead of a tracker for 3D-3D registration of the surgical scene with the pre-operative model data. An example for 2D-3D registration is presented in [15], where navigation is simplified by displaying the corresponding pre-operative CT or MRI slice next to the intra-operative endoscopic image.

2 Methods

In this section the experimental setup and the algorithms used for the 3D reconstruction based on the point database system and the 3D-3D registration between the computed model and the pre-operative data are presented.

2.1 Experimental Setup

The entire hardware setup is depicted in Fig. 1. For our experiments a 10 mm radial distortion corrected endoscope (Richard Wolf GmbH) with an oblique viewing angle of 25° was used. To avoid interlacing artifacts, we relied on a progressive frame color CCD camera with a resolution of 960×800 pixels and 15 fps. As the depth-of-field of the endoscope/camera combination is in the range of $3 - 8$ cm the focal length of the camera can be kept constant during the entire procedure, allowing to avoid the recalibration of the system during surgery.

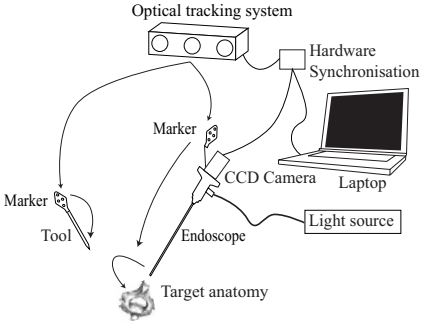


Fig. 1. Overview of the different components of the endoscopic navigation system

A marker is attached to the endoscope that is being followed by the active optical tracker (EasyTrack500, Atracsys LLC). The EasyTrack provides accurate position (less than 0.2 mm error) and orientation information in a working volume of roughly $50 \times 50 \times 1500\,\mathrm{cm}^3$. An external hardware triggering logic ensures the synchronized acquisition of the tracking data and the camera images during dynamic freehand manipulation.

The entire setup is portable and can be installed in a sterile OR environment within minutes. The intrinsic camera and the extrinsic camera-marker calibration can be performed pre-operatively in one step by the surgeon without requiring assistance from a technician [16].

2.2 Point Database

Efficient and consistent data management is an important pre-requisite for the proposed near real-time 3D reconstruction method. All the captured views with their interest points, the views where a specific region is visible, and the continuously updated 3D point coordinates have to be stored and managed reliably and efficiently. In this section a novel approach for handling these data is presented. The algorithms described in the following section rely fully on this database.

The data structure to efficiently handle the related task is shown in Fig. 2 and consists of three parts. First, all the images with the camera pose, the location of the interest points and their feature descriptors are stored in the `Interest`

`Point Database`. Second, in the tracking phase the point correspondences between the last two views are established. If a feature was already detected in more than two views, the actual view is added to this feature. If the feature has only been present in the last two views, the newly matched points are added to the `Tracking Database`. Finally, during 3D reconstruction and model refinement the `3D Point Database` is filled by triangulating the matched image points from the `Tracking Database`. This database contains the 3D coordinates, the views these points were visible in and the locations of the corresponding 2D interest points. In order to maximize speed and minimize memory load, all references are realized using pointers.

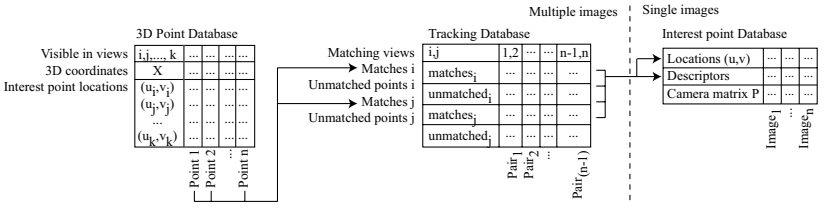


Fig. 2. Database structure for tracking, reconstruction and model refinement

2.3 Algorithm Outline

The method can be divided into three parts: (1) the tracking of feature points across the image sequence, (2) the 3D reconstruction and (3) the registration or referencing of the computed 3D structure.

As the endoscope rarely sees the whole scene at once, we need to find natural landmarks having distinctive features that can be tracked across multiple images and found again when reappearing. Figure 3 gives an overview of the steps performed in the feature tracking algorithm. To initialize the tracking two images are required. Starting with the first image, additional images are acquired until a baseline larger than the predefined threshold of 4 mm is detected. The baseline can be easily calculated using the extrinsic parameters reported by the external tracker. This minimum baseline restriction assures an accurate initial 3D reconstruction. From these two views interest points and their region descriptors are extracted using the method described in [17]. They are matched using the second-nearest-neighbor ratio of their 64-element feature descriptors. A filter using the robust RANSAC fundamental matrix estimation is then used to remove outliers. The remaining point correspondences are then used to triangulate an initial set of 3D points which are stored in the `3D Point Database`.

Once the point tracking is initialized, the trifocal tensor is calculated for the already known 3D points with the last two camera poses where the point was visible. The trifocal tensor is then used to predict the 2D positions of the 3D points in the current endoscopic view. If a point is indeed visible in the

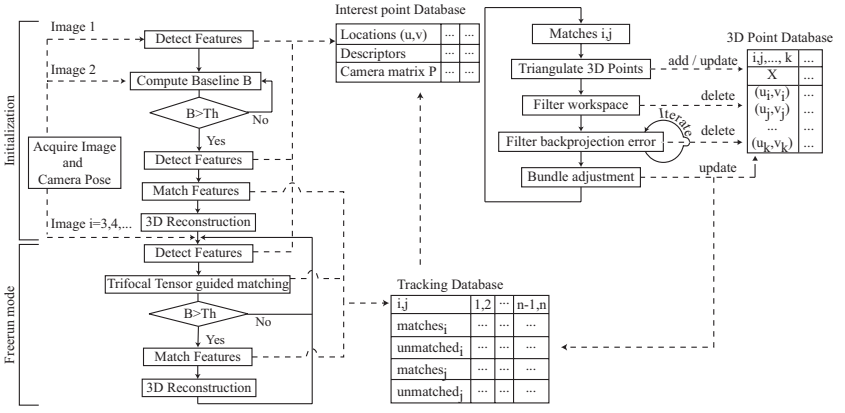


Fig. 3. Tracking and 3D reconstruction algorithm

current view, this view is added to the corresponding `3D Point Database` entry. If the baseline between the current view i and the last view $i-1$ is less than the predefined threshold, the algorithm captures a new image and starts over with calculating the trifocal tensors. Otherwise, the still unmatched interest points in the current view i are checked if they can be matched to an interest point in the previous view $i-1$. These new points are added to the `Tracking Database` and the triangulated 3D coordinates are stored in the `3D Point Database`.

In the feature point tracking phase (see Fig. 3 (left)) the `3D Point Database` is continuously updated. In the 3D reconstruction phase (right) the newly found 3D points are triangulated using the optimal triangulation method [18] that shifts the image coordinates to perfectly fulfill the epipolar constraint. The computed 3D points are then added to the `3D Point Database`. As we need accurate 3D points for further processing, two different filters are used to remove false results. The first filter eliminates all 3D points from the database, that are outside the working volume of the external tracking device. The second filter computes the backprojection error for each point in the last 5 views. Iteratively all points having a backprojection error larger than the $threshold = median + 3 \cdot \sigma$ are removed. The median and standard deviation are then recalculated for the remaining points. This filter stops once no more points have been eliminated from the database (usually after two iterations). After the filtering, the database contains a stable set of 3D points. The Levenberg-Marquardt (LM) algorithm is then used to optimize the structure parameters, over the last 5 views. The optimized 3D points are then updated in the `3D Point Database`.

In order to provide a good initialization for the ICP registration, the presegmented CT model is manually overlaid on the point cloud. Therefore the point cloud needs to be dense enough for identifying common structures of the two models.

3 Results

In order to assess and quantify the system's performance, we tested the algorithms on synthetic data (C2 vertebra phantom, mimicking blood spots and coagulation residuals) and on ex-vivo data. Three different accuracy tests and a feasibility study based on cadaver experiments were performed using only freehand captured images.

In the first experiment the quality of the metric 3D reconstruction was determined. Therefore, the 3D structure from 25 different sequences of a plastic vertebra were computed. The real thickness of the spinal process of the C2 was then compared to the reconstructed models. The obtained error was 0.10 ± 0.35 mm with the maximum error smaller than 0.8 mm. An example of the reconstructed C2 vertebra and the corresponding CT model is shown on Fig. 4a, b and c.

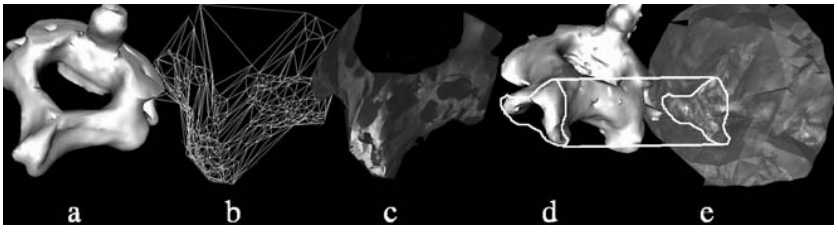


Fig. 4. a) Phantom CT model, b) wireframe and c) textured model of reconstructed phantom, d) Cadaver C2 CT model and e) corresponding textured reconstruction

In the second experiment the repeatability of the localization has been investigated. Therefore, the vertebra was fixed and 10 different sequences recorded followed by the 3D reconstruction. Each of the resulting models was then registered to the CT data using the point-to-point ICP. Then the spatial transformation between the registered models was measured, resulting in a deviation of 0.46 ± 0.31 mm in the position of the vertebra and $1.3 \pm 0.8°$ for the rotation.

In the third experiment the precision of the quantified displacements has been assessed. The vertebra was positioned at different locations using a robot arm. At every distinct position the vertebra was reconstructed in 3D and registered to the CT . The distance between the registered CT models was then measured and compared to the ground truth from the robot thus covering the full error chain. This test has been performed separately in the xy-plane of the tracker's coordinate system with an error of 0.42 ± 0.28 mm as well as in its z-direction yielding an error of 0.49 ± 0.35 mm.

In order to verify the performance of the 3D reconstruction in a more realistic environment, we tested it on ex-vivo data from a cadaver experiment. On average we could identify 500 points on the cadaver's C2 vertebra, which proved to be sufficiently dense for the 3D reconstruction and compares well with the 800 points found on the synthetic body. We observed this similarity in the backprojection error, too, which was (2 ± 1 pixels) in both cases, supporting our expectations to be able to reach similar accuracy as in the synthetic environment. Fig. 4d and

4e show the reconstructed pre-operative CT and intra-operative optical models, enabling a visually satisfactory overlay. These preliminary results indicate the usefulness of our method in an in vivo surgical environment.

The feature tracking and consecutive 3D reconstruction process currently uses approx. 1 s per frame. The final ICP registration step uses slightly less than 1 s, thus allowing the surgeon to reference the target anatomy in near real-time.

For the proposed system the following sources of inaccuracy have been identified: (1) intrinsic- and extrinsic errors in camera calibration, (2) inaccuracy of the external tracking device, (3) inaccuracies when matching features, and finally (4) errors from the ICP registration. The potential for improvements in intrinsic and extrinsic calibration is very limited, so further efforts for achieving even higher accuracy will have to concentrate on the last three components.

4 Conclusions

In this paper, a purely optical natural landmark based registration and referencing method using a tracked monocular endoscope was presented. The method is non-invasive and no fiducials or other synthetic landmarks are required. The metric 3D reconstruction accuracy is 0.1 mm. Rigidly registering the 3D reconstruction to the pre-operative 3D CT model using point-to-point ICP resulted in an error of < 0.5 mm proving the applicability of the proposed procedure for non-invasive registration or referencing during navigated spine surgery. Compared to similar work, a high accuracy is achieved, although it is difficult to compare these results due to the different setups and the lack of standardized benchmarks. The need for texture and distinguishable structure on the object for the registration are the two main short-comings of the algorithm. Currently motion in the scene cannot be handled, but it will be investigated to integrate parametrized motion models.

The current processing speed of one second per frame allows the surgeon to register the target anatomy within about thirty seconds. The current implementation, however, still leaves room for further improvement in processing speed.

As the next step, the cadaver experiments will be extended allowing quantitative comparisons with external tracking results, followed by in vivo validation during open C1/C2 transarticular screw placement surgery.

Acknowledgments

This work has been supported by the CO-ME/NCCR research network of the Swiss National Science Foundation (http://co-me.ch).

References

1. Schlenzka, D., Laine, T., Lund, T.: Computer-assisted spine surgery. European Spine Journal **9 (Suppl 1)** (2000) 57–64
2. Kowal, J., Amstutz, C., Ioppolo, J., Nolte, L.P., Styner, M.: Fast automatic bone contour extraction in ultrasound images for intraoperative registration. IEEE Trans. Med. Imag. (2003)

3. Dey, D., Gobbi, D., Slomka, P., Surry, K., Peters, T.: Automatic fusion of freehand endoscopic brain images to three-dimensional surfaces: creating stereoscopic panoramas. IEEE Trans. Med. Imag. (2002)
4. Thoranaghatte, R.U., Zheng, G., Langlotz, F., Nolte, L.P.: Endoscope based hybrid-navigation system for minimally invasive ventral-spine surgeries. Computer Aided Surgery (2005) 351–356
5. Mori, K., Deguchi, D., Akiyama, K., Kitasaka, T., Maurer, C.R., Suenaga, Y., Takabatake, H., Mori, M., Natori, H.: Hybrid bronchoscope tracking using a magnetic tracking sensor and image registration. In: MICCAI. (2005) 543–550
6. Burschka, D., Li, M., Ishii, M., Taylor, R.H., Hager, G.D.: Scale-invariant registration of monocular endoscopic images to ct-scans for sinus surgery. Medical Image Analysis **9** (2005) 413–426
7. Koppel, D., Wang, Y.F.: Image-based rendering and modeling in video-endoscopy. In: Proc. of the IEEE International Symposium on Biomedical Imaging: Macro to Nano. Volume 1. (2004) 269 – 272
8. Caban, J.J., Seales, W.B.: Reconstruction and Enhancement in Monocular Laparoscopic Imagery. In: Proc. of Medicine Meets Virtual Reality. Volume 12. (2004)
9. Feuerstein, M., Wildhirt, S.M., Bauernschmitt, R., Navab, N.: Automatic Patient Registration for Port Placement in Minimally Invasive Endoscopic Surgery. In: MICCAI. (2005)
10. Bockholt, G., Bisler, U., Becker, A.: Augmented reality for enhancement of endoscopic interventions. In: Virtual Reality Proceedings. (2003)
11. Deligianni, F., Chung, A., Yang, G.Z.: Visual Augmentation for Virtual Environments in Surgical Training. In: MICCAI. (2003)
12. M. Scholz, e.a.: Development of an Endoscopic Navigation System Based on Digital Image Processing. Computer Aided Surgery **3** (1998)
13. Sauer, F., Khamene, A., Vogt, S.: An Augmented Reality Navigation System with a Single-Camera Tracker: System Design and Needle Biopsy Phantom Trial. In: MICCAI. (2002) 116–124
14. Mourgues, F., Devemay, F., Coste-Maniere, E.: 3D reconstruction of the operating field for image overlay in 3D-endoscopic surgery. In: Proc. of the IEEE and ACM International Symposium on Augmented Reality. (2001)
15. Muacevic, A., Muller, A.: Image-Guided Endoscopic Ventriculostomy with a New Frameless Armless Neuronavigation System. Computer Aided Surgery **4** (1999)
16. Wengert, C., Reeff, M., Cattin, P., Székely, G.: Fully automatic endoscope calibration for intraoperative use. In: BVM, Springer-Verlag (2006) 419–23
17. Bay, H., Tuytelaars, T., Van Gool, L.: Surf: Speeded up robust features. In: Proc. of the ninth European Conference on Computer Vision. (2006)
18. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press (2000)

Real-Time Tracking of Contrast Bolus Propagation in Continuously Moving Table MR Angiography

Joshua Trzasko[1], Stephen Riederer[1,2], and Armando Manduca[1,3]

[1] Mayo Clinic College of Medicine, Rochester, MN 55905
[2] Magnetic Resonance Research Lab, Mayo Clinic, Rochester, MN 55905
[3] Biomathematics Resource, Mayo Clinic, Rochester, MN 55905

Abstract. The recent introduction of variable-velocity control for Continuously Moving Table Magnetic Resonance Imaging has introduced the ability for interactive examination across extended fields of view. A common application of this method is contrast-enhanced angiography, where propagation of an intravenously injected contrast agent is tracked throughout the peripheral vascular tree. Whereas current methods for performing contrast bolus tracking are entirely manual, we discuss the complete automation of this process through the use of real-time image processing. Specifically, we provide a coupled intensity-correction procedure and modified Fast Marching method for rapid segmentation of contrast enhanced vasculature in CE-MRA and discuss the incorporation of this process into a framework for fully automated and adaptive control of table motion for real-time tracking of contrast bolus propagation through the lower peripheral vasculature.

1 Introduction

Contrast-enhanced MR Angiography in its present form was first proposed by Prince et al. [1] in 1993 and has since become a dominant modality in the diagnosis of vascular disease. One common application of CE-MRA is in the diagnosis of lower peripheral vascular disease (PDV), a condition where atherosclerotic build-up leads to tissue ischemia in the extremities. As the spatial extent of the lower peripheral vasculature significantly exceeds the maximal field-of-view (FOV) that can be imaged in a single acquisition by contemporary MRI systems, it is inevitably necessary to move the patient table during the examination to analyze the entire region of interest. Movement of the patient bed to allow full coverage of the extended FOV may be performed either during the intermission between stationary imaging at fixed stations or continuously during the image acquisition process. The main challenge faced in extended FOV imaging is providing optimal arterial phase acquisition with minimal venous contamination at all imaged positions [2].

The current standard clinical protocol for this type of exam uses the fixed-station approach, with stationary images being acquired under steady-state contrast concentration conditions [3]. Patient-specific flow information is not taken into account to yield optimal arterial-phase images and, resultantly, there is a propensity for venous contamination to occur in images acquired at inferior stations. Furthermore, the steady-state nature of images acquired under this paradigm cannot provide information about contrast dynamics for use in flow-based diagnostics.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 824 – 831, 2006.

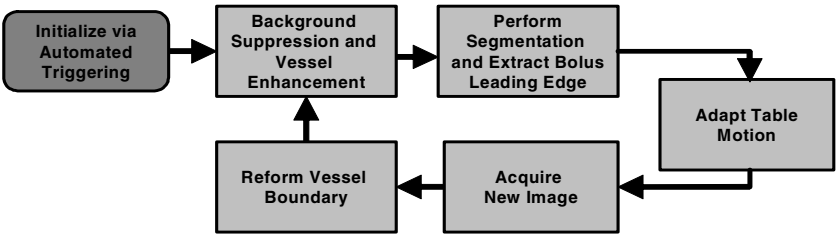


Fig. 1. Outline of Automated Bolus Tracking System

The use of a continuously moving table (CMT) during image acquisition overcomes several of the shortcomings of the fixed-station approach to extended FOV imaging [4,5]. Under this formulation, a single image depicting the entire extended FOV is acquired progressively and displayed in real-time through the use of fast reconstruction hardware [6]. Consequently, the velocity of the patient bed can be interactively controlled by the scanner operator in response to variation of contrast bolus dynamics at different regions of the body. This ability guarantees optimized arterial phase acquisition at all positions within the extended field-of-view. While bolus flow information is not assessed quantitatively when using this method, the ability to track the leading edge of the contrast bolus inherently provides a framework under which such information can be measured.

We propose a novel paradigm for fully-automated tracking of contrast bolus propagation via real-time segmentation of the contrast-filled vasculature and discuss how this process may be embedded within the CMT CE-MRA framework to provide an optimal arterial-phase, patient-specific, and autonomous image acquisition procedure. In addition to reducing operator burden and associated potential error, the progressive assessment of contrast flow resulting from the segmentation process can potentially be used for diagnostic analysis of flow-related diseases such as diabetes.

2 Methods

The procedure for tracking the propagation of the contrast bolus can divided into a series of distinct processes, as shown in Figure 1. We first discuss the initialization of the tracking procedure using an automated triggering process and then develop an intensity remapping technique which enhances vascular signal prior to segmentation. A modification to the well-known Fast Marching method for curve evolution is then presented for optimized segmentation of the contrast-enhanced vasculature under the CMT framework, and we address the process utilizing the segmentation results for adaptive control of table motion dynamics.

2.1 Initialization Via Automated Triggering

Initialization of bolus-chase CE-MRA studies of the lower peripheral vasculature typically occurs following observation of contrast influx into the thoracic aorta. It has been previously shown that these studies may be automatically triggered using

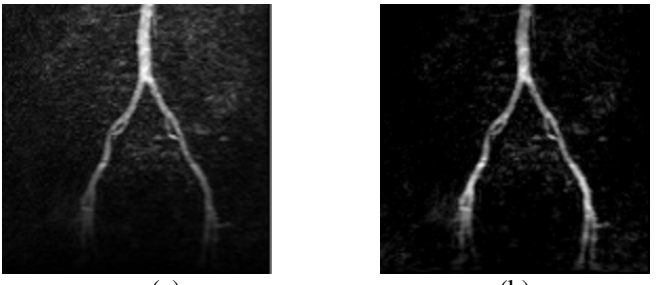
real-time image processing techniques for sequential detection and analysis of key physiological events that occur prior to aortic filling [7]. We employ this recognition process for automatic placement of an initial seed point for Fast Marching-based segmentation.

2.2 Background Suppression and Enhancement of the Bolus Leading Edge

In CE-MRA, signal can generally be assessed as belonging to one of two classes: that arising from intraluminal contrast agent presence and that originating from within the background. The transient nature of intravascular contrast concentrations at different positions within the visible FOV, however, leads to strong intensity variation within both signal classes and thus a global intensity-based discriminant cannot be defined. For example, moderate soft-tissue enhancement from interstitial contrast diffusion tends to occur adjacent to vascular regions that have already reached steady-state contrast concentration levels. Despite the elevated background signal intensity, high contrast concentrations within the vasculature still provide adequate distinction between the two signal classes although the dividing threshold will be high. Conversely, tissues surrounding transient-state intravascular contrast signal will not be enhanced and thus signal arising from these regions will be minimal. Notwithstanding low intraluminal signal, adequate distinction between the signal classes is still present albeit the dividing threshold is significantly lower than in the steady-state case.

As a single threshold is requisite for termination of any Fast Marching scheme, we perform an intensity remapping across the image such that the local thresholds dividing the signal classes, as determined by intravascular contrast concentration, are all mapped to a constant and predetermined value. In continuously moving table CE-MRA, both the direction of contrast flow and the direction of table motion lie predominantly along the superior-inferior (SI) direction. It can be inferred that intravascular contrast concentration and ultimately the intensity threshold dividing the signal classes will also be a function of SI position. A uniform threshold can be achieved by scaling all points associated with a particular SI position such that the class discriminant is equal to the desired threshold.

As intravascular contrast signal is sparse within an image, it is statistically more powerful to estimate the termination threshold as an upper bound on the background signal class. The use of standard moments as estimators of the background class parameterization, however, will create a positively biased model resultant from the inclusion of high intensity intravascular contrast signals and the derived upper bound will lie significantly higher than that of the background class. Attempts to normalize the background across the image will result in equalization of contrast signal and background signal on lines containing no contrast presence. Resultantly, curve propagation will not be terminated prior to exiting the vascular region. The adverse effects associated with positive bias introduced from intravascular signal when using standard statistical moments can be avoided by means of robust estimators which are immune to outlier influence. A Gaussian model can be utilized as an asymptotic



(a) (b)

Fig. 2. CE-MRA image before (a) and after (b) intensity correction

upper bound to the Rician density [8] that governs noise in magnitude MR images. The mean and variance of the background signal class for an image I, taken as a function of SI position, y, can be estimated as

$$\mu(y) \approx \underset{\forall x, y' \in [y-\delta, y+\delta]}{median} \left(I(x, y') \right)$$
$$\sigma^2(y) \approx \left[\frac{\underset{\forall x, y' \in [y-\delta, y+\delta]}{MAD} \left(I(x, y') \right)}{\Phi^{-1}\left(3/4\right)} \right]^2, \quad (1)$$

where MAD is the median absolute deviation operator and Φ^{-1} is the quantile function. Non-zero parameterization of the variable δ allows the estimation process to be performed over a local window surrounding the SI position of interest, offering increased invariance to contrast-based signal and thus an improved measure of the background noise process. The value $\Phi^{-1}(3/4)$ arises as the expected value of *median*($|Z|$) for a random variable $Z \sim N(0,1)$, and its use as a scaling factor in (1) forces the MAD operator to be an unbiased estimator of the standard deviation [9]. Using the estimates in (1), an image can then be mapped via

$$I'(x, y) = \begin{cases} \dfrac{I(x, y) - \mu(y)}{\sigma(y)} & I(x, y) \geq \mu(y) \\ 0 & else \end{cases}, \quad (2)$$

and the uniform termination threshold for the Fast Marching scheme will be the z-score associated with the desired confidence of background signal containment, e.g. $z = 1.65$ will provide a bound which lies above 95% of the background signal class. As the upper bound on the background signal will always be above the mean of the class for a given SI position, any signal below that value can be neglected and set to zero with no effect on the curve evolution process. This action preserves the non-negativity property of the speed function when it is defined in terms of the intensity map. An example of the mapping results described in (2) can be seen in Figure 2.

2.3 Real-Time Vessel Segmentation Using a Modified Fast Marching Method

The Fast Marching method [10,11] is a non-iterative upwind technique for solving the Eikonal equation

$$\left|\nabla T\right| F = 1, \tag{3}$$

where T is the arrival time function and F is a strictly non-negative speed function. Whereas F is generally defined as a function of the image gradient and (3) is subsequently minimized, the dispersive nature of the contrast bolus at the leading edge does not provide an adequate boundary to restrict the evolution of the front to only within the vasculature and thus this approach is not valid for our particular application. Following the remapping procedure discussed in the previous section, we have a bimodal intensity distribution with a known discriminant threshold. We address the problem associated with gradient-based control of the Fast Marching curve evolution process by defining our speed function, F, to be a function of remapped image intensity, specifically

$$F = G * I', \tag{4}$$

where G is a smoothing kernel, and we instead maximize the relation in (3). Under this formulation, curve evolution will be driven towards image regions containing maximal contrast presence. The uniform termination threshold for I' will allow curve propagation into transient contrast concentration states such as the leading edge of the bolus but will restrict entrance into the non-vascular background despite elevated intensity due to soft-tissue enhancement. Following segmentation, determination of the leading edge of the bolus becomes a trivial task.

While implementation of the Fast Marching segmentation process using a traditional gradient-based speed term is traditionally performed using a min-heap storage structure for minimization of (3), as we are instead maximizing the Eikonal equation in our implementation, optimal execution of our segmentation process can be achieved though use of a max-heap storage structure. Additionally, we can further optimize the segmentation process for real-time implementation by embedding the smoothing process directly into the curve evolution procedure. In applications involving evolutionary modeling of a growing object across a series of temporally-varying frames, the inter-frame variability in conformation of the front tends to be minimal and locally restricted relative to the preceding bounds. Despite these observations, subsequent frames are typically filtered globally and many of the smoothed points are never utilized. Prior to insertion of an image point onto the heap structure representing the segmentation curve, a single-element convolution can be executed. By restricting the domain of convolved pixels to consist of only points with which the front interacts during its evolution, the computational complexity of the smoothing process is inherently reduced from $O(N^2)$ to $O(1)$ while yielding the same effective results as if global smoothing had been performed.

2.4 Control of Table Motion Dynamics

As the velocity of blood and consequentially the contrast bolus varies significantly between different vascular regions, the speed of table motion must also vary if the

bolus leading edge is to be kept within the visible field of view at all times. Knowledge of the current and past bolus leading edge positions can be used as the basis for feedback control of table motion such that the predicted future position of the bolus leading edge will reside within the visible field of view. Implementation of such a control system is under way and will be discussed in a later work.

2.5 Inter-frame Curve Evolution

Between subsequent image frames, the magnitude of the positional change of the leading edge of the bolus will far exceed that due to local variations arising from pulsatility. Consequentially, contrast propagation will be monotonically outward and thus any prior segmentation boundary can be used as the initial segmentation boundary within a newly acquired frame. When using the boundary from the frame acquired immediately before the newly acquired image, expansion of the curve is limited to regions where contrast presence is new to the scene, as all steady-state regions will have been previously segmented. The effect of this inter-frame propagation approach is that, for any newly acquired image, segmentation of the vasculature can be achieved in the minimal possible number of operations for that particular frame.

Prior to allowance of a past segmentation boundary as the initial condition to the segmentation procedure for a newly acquired image frame, the boundary must first be updated to account for change of the visible field of view as well as changes in the intensity values associated with a particular spatial position. A shift in the visible field of view results from table motion that occurs during the bolus tracking procedure, and all boundary elements which have exited the scene must be removed as they are no longer valid for further evolution. The physical shift between sequential image frames can be obtained directly from the imaging system, and thus pruning becomes essentially trivial. Changes in image intensity between frames for a given spatial position arise from flow of the contrast into new vascular regions. All remaining boundary points visible within the new scene are placed onto a new heap structure and segmentation resumes. As the heap sort procedure is $O(N\log N)$, the boundary reformation process can be achieved quickly.

3 Experimentation and Results

All images used in experimentation were acquired on a General Electric Signal 1.5T MRI system using a time-resolved gradient echo (GRE) pulse sequence [12] and reconstructed using the continuously moving table methods described by Kruger et al. [4,5]. Images were typically acquired with the dimensions (X = Y = 128, Z = 8) and the prescribed visible fields of view were (FOV_X = FOV_Y = 30-40cm, FOV_Z = 10cm; under coronal orientation, X corresponds to the Superior/Inferior axis, Y corresponds to the Medial/Lateral axis, and Z corresponds to the Anterior/Posterior axis. The temporal resolution of image acquisition was 0.6s per frame. Following volume acquisition, a maximum intensity projection (MIP) was computed along the Z direction to produce the 2D images used in all processing methods described in this paper. Real-time processing of image data was performed using the Java programming platform on a 3GHz Pentium 4 PC containing 1GB RAM. The average processing time

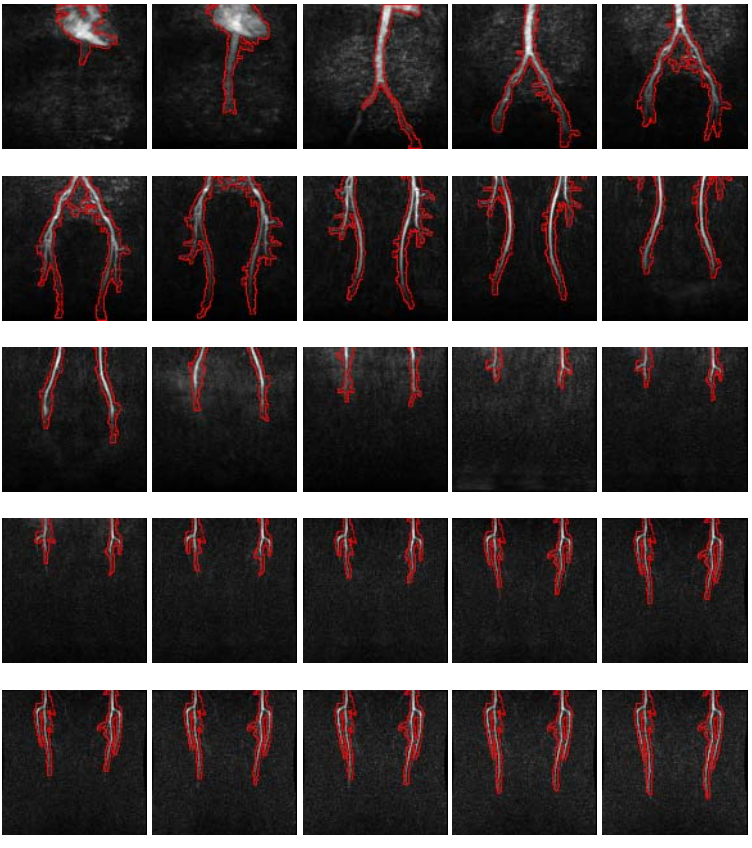


Fig. 3. Tracking of contrast bolus propagation across the lower peripheral vasculature in a sample data set

per frame was approximately 25 ms and thus automatic table control is fully realizable within the real-time acquisition framework.

The automated bolus tracking method has been initially tested on six clinical extended field of view studies acquired under constant table motion (3-4 cm/s); extension to variable velocity testing will follow completion of the control system discussed in section 2.4. Figure 3 shows a sample series of images from one exam along with the associated segmentation produced by our algorithm. As is evident in the depicted example, the tracking results clearly correlate with the visible leading edge of the contrast bolus at all spatial positions within the extended field of view and the other five tested cases gave similar results.

4 Conclusion

We have presented a framework for autonomous tracking of contrast bolus propagation across and extended FOV using real-time CE-MRA. Through the incorporation

of real-time image processing techniques into the acquisition framework, optimal arterial phase images can be acquired, operator burden can be significantly alleviated, and both anatomical and physiological flow information are simultaneously measured. Potential future applications include the mapping of contrast bolus velocities onto the anatomical map of the vasculature tree produced by the segmentation process for use in detection of turbulent flow, the latter which has been linked to the formation of atherogenic plaques [13]. Other similar applications are also possible.

References

1. M. Prince, E. Yucel, J. Kaufman, D. Harrison, and S. Geller. Dynamic Gadolinium-enhanced Three-dimensional Abdominal MR Arteriography. *JMRI*, 3: 877-881, 1993.
2. J. Maki, M. Prince, F. Londy, and T. Chenevert. The effects of time varying intravascular signal intensity and k-space acquisition order on three-dimensional MR angiography image quality. *JMRI*, 6(4): 642-651, 1996.
3. K. Ho, T. Leiner, M. de Haan, A. Kessels, P. Kitslaar, and J. van Engelshoven. Peripheral vascular tree stenoses: evaluation with moving-bed infusion- tracking MR angiography. *Radiology*, 206: 683-692, 1998.
4. D. Kruger, S. Riederer, R. Grimm, and P. Rossman. Continuously Moving Table Data Acquisition Method for Long FOV Contrast-Enhanced MRA and Whole-body MRI. *MRM*, 47: 224-231, 2002.
5. D. Kruger, S. Riederer, J. Polzin, A. Madhuranthakam, H. Hu, and J. Glockener. Dual-Velocity Continuously Moving Table Acquisition for Contrast-Enhanced Peripheral Magnetic Resonance Angiography. *MRM*, 53: 110-117, 2005.
6. S. Riederer, T. Tasciyan, F. Farzaneh, J. Lee, R. Wright, and R. Herfkens. MR Fluoroscopy: Technical Feasibility. *MRM*, 8: 1-15, 1988.
7. J. Trzasko, A. Madhuranthakam, D. Kruger, and S. Riederer. Real-Time Triggering and Tracking of the Contrast Bolus Leading Edge in Time-Resolved Continuously Moving Table CE-MRA. *Proceedings of the 13th Scientific Meeting of the ISMRM*, 2005.
8. S. Rice. Mathematical Analysis of Random Noise. *Bell System Technical Journal*, 23: 282-332, 1944; 24: 46-156, 1945.
9. P. Huber. Robust Statistics. Wiley: New York, 1981.
10. J. Sethian. A Fast Marching Level Set Method for Monotonically Advancing Fronts. *PNAS*, 93: 1591-1595, 1996.
11. R. Malladi and J. Sethian. An O(NlogN) Algorithm for Shape Modeling. *PNAS*, 93: 9389-9392, 1996.
12. A. Madhuranthakam, D. Kruger, S. Riederer, J. Glockner, and H. Hu. Time-Resolved 3D Contrast-Enhanced MRA of an Extended FOV Using Continuous Table Motion. *MRM*, 51: 568-576, 2004.
13. Y. Fung. Biomechanics: Circulation. Springer-Verlag: New York, 1997.

Preventing Signal Degradation During Elastic Matching of Noisy DCE-MR Eye Images

Kishore Mosaliganti[1], Guang Jia[2], Johannes Heverhagen[2], Raghu Machiraju[1], Joel Saltz[3], and Michael Knopp[2]

[1] Department of Computer Science and Engineering
[2] Radiology
[3] Biomedical Informatics, The Ohio State University, Columbus, OH, USA
kishore@bmi.osu.edu

Abstract. Motion during the acquisition of dynamic contrast enhanced MRI can cause model-fitting errors requiring co-registration. Clinical implementations use a pharmacokinetic model to determine lesion parameters from the contrast passage. The input to the model is the time-intensity plot from a region of interest (ROI) covering the lesion extent. Motion correction meanwhile involves interpolation and smoothing operations thereby affecting the time-intensity plots. This paper explores the trade-offs in applying an elastic matching procedure on the lesion detection and proposes enhancements. The method of choice is the 3D realization of the Demon's elastic matching procedure. We validate our enhancements using synthesized deformation of stationary datasets that also serve as ground-truth. The framework is tested on 42 human eye datasets. Hence, we show that motion correction is beneficial in improving the model-fit and yet needs enhancements to correct for the intensity reductions during parameter estimation.

1 Introduction

The model-based quantitative analysis of contrast agent diffusion in dynamic contrast enhanced MRI (DCE-MRI) allows us to estimate the magnitude and spatial distribution of physiological kinetic parameters such as amplitude (A), redistribution rate constant (k_{ep}) and the elimination rate constant (k_{el}) [1]. The quantification of the temporal intensity signal in the tumor region reflects its angiogenic properties such as the vascular density, arrangement and functional permeability within [2]. The input to the model is a sequence of DCE-MR images from which a region-of-interest (ROI) encompassing an entire or portion of the lesion is selected. Quantification involves fitting an appropriate curve to the temporal intensity pattern drawn from the ROI. The modeled parameters can be used to relate to lesion specific physiological concepts.

The challenging aspect in quantification is the positioning of the ROI consistently to cover a lesion across all the epochs. The time required for imaging, non-stationary organs such as the eye, kidney, liver and lung, patient movement, breathing, pain and gravitation effects can cause elastic deformation [2]. Hence,

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 832–839, 2006.

model fitting errors arise that alter voxel-to-tissue mapping necessitating the use of multiple ROI's across the temporal scale.

Motion correction has been proposed in the past for the DCE-MRI sequences for improving the parameter estimation by placing a single ROI that is temporally consistent. The choice of using rigid [3, 4, 5] or elastic [6, 7] matching methods have depended on the organ and the deformation inherent in it. While initial research focussed on using conventional metrics (e.g. mutual information [8], correlation ratio [9]) with limited success, efforts were made to incorporate the quantification model into the registration cost function [3]. Recently, Buonaccorsi et al. [5] proposed using a simplified tracer model directly into the registration metric to develop a localized, 3D translational registration algorithm focussing on the tumor VOI to obtain better model fits. They assumed that the abdominal tumors were sufficiently rigid in comparison to surrounding tissue. In our case, the low spatial resolution of the eye images (256x256) with high acquisition noise does not provide satisfactory rigid registration results. The tumors in the eye deform elastically due to subject movement, soft tissue and eye ball rotation.

While the need for motion correction has been reported in the literature, the presence of registration artifacts has not been investigated. Co-registration is concomitant with image transformations leading to interpolation of the image field on a grid. Interpolation is akin to a low-pass filtering operation and hence pixel intensities are affected. Reduction in pixel intensities affect the amplitude, rise and wash-in or wash-out gradients of the time-intensity curves and hence the parameters.

In this paper, we explore the role of the Demons registration method [10] in providing a better model-fit. Our research leads us to evaluating the intensity changes caused by smoothing during the registration procedure. Experiments are performed using stationary human eye datasets with induced elastic motion to explore signal degradation. We perform *anisotropic smoothing* to reduce high frequency noise prior to a *high-pass filtering* stage. The high-pass filter compensates the smoothing that occurs during registration. Finally, *histogram matching* of the registered image to the original image is done to match the intensities. The framework is then applied to 42 clinical datasets having complex motion. We show that the pharmacokinetic model provides a better fit to the registered signal patterns as compared to the unregistered case. Moreover, the fit is comparable to the case where manual ROI placement is done across all time-points. We show examples where exact boundaries of the tumor are better identified owing to the improved registration. Hence, we motivate the need for applying motion-correction with smoothing compensation into the quantification of DCE-MRI.

2 3D Elastic Co-registration: Thirion's Demons

Thirion [11, 12] introduced the concept of diffusing models to perform image matching, an essential component of this study. Image matching is performed through the movement of a deformable grid through a semi-permeable contour of an object surface in the other image. The method was validated using synthesized

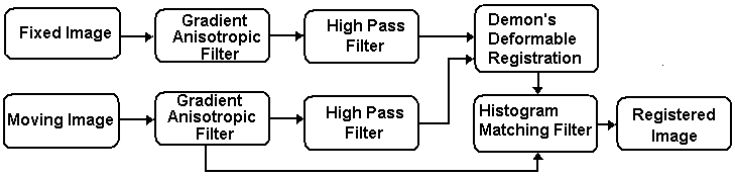


Fig. 1. Deformable registration framework with intensity compensation

deformation on medical images. The method was applied to track heart-motion and 3D matching of inter-patients brain images [13]. In [10], a study of evaluating the temporal variations of lesion volumes for practical applications such as therapeutic intervention effects, therapy monitoring and pharmaceutical trials was conducted. This places our work in perspective. For the DCE-MRI images, each of the epochs provides a $3D$ volume which is registered to a pre-determined base 3D volume, chosen at the peak contrast when the features are best defined. The demon's method (see Figure 1) was implemented using the National Library of Medicine NLM/NIH's Insight Toolkit (ITK) [14].

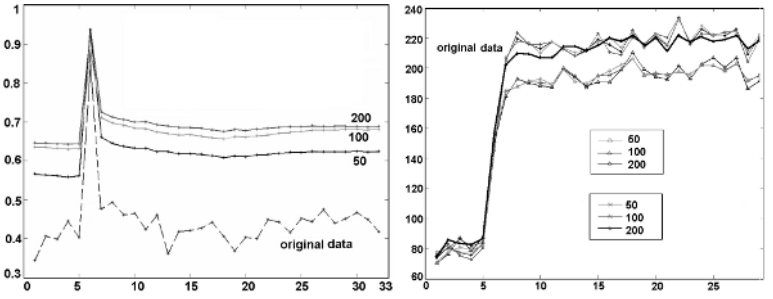
A total of 50 iterations yielded satisfactory results with the standard deviation of the Gaussian kernel smoothing the deformation field set at 1.0. The smoothing is essential to avoid inconsistent deformations from occurring. The other components of the framework, namely the gradient-anisotropic filter, high-pass filter and histogram matching filter are explained in the following section.

3 Validation of Intensity Compensation

The pharmacokinetic model estimation of the lesion parameters depends on the temporal intensity variation in the ROI encompassing the tumor. Earlier, we mentioned that registration causes smoothing of the image due to the interpolation of the warped image onto a grid. We now seek to understand registration performance based on the changes in pixel intensities. Our task is complicated by the fact that there is no ground truth for pixel intensity validation after registration since a 1-to-1 correspondence with the original image does not exist. Towards this purpose, datasets with no motion in them are selected. Motion is artificially induced in such datasets and then registered. The registered dataset is then compared to the original to study the smoothing effects.

Forty-two patients with eye tumors were included in the study protocol. DCE-MRI datasets were acquired on a clinical 1.5-T MR system (GE SIGNA) using a fast gradient-echo sequence ($3D$-FSPGR). Total scan time per volume was about 8 minutes. After the third phase, a small molecular weight paramagnetic contrast agent (e.g. Gd-DTPA, Magnevist) was injected using a power injector at a constant infusion rate of approximately $18s$. $3D$ volumes of the eye at 28 time-points were acquired, with peak contrast being evident in the 6^{th} epoch.

After manually scanning the forty-two datasets for stationary datasets, EYE-OS-P05 and EYE-OS-P28 were selected for this part of the study. We synthesize



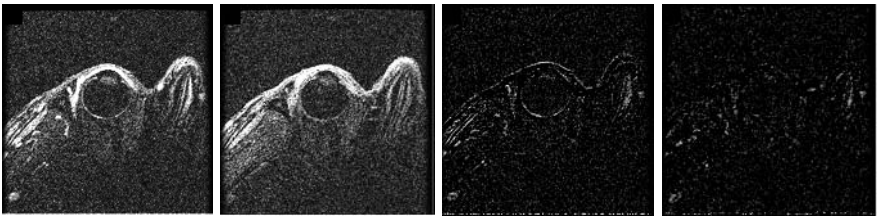
(a) (b)

Fig. 2. (a): Pearson's correlation is plotted for each of the epochs before and after registration. The 6th epoch is chosen as the base volume causing the occurrence of a peak. (b): The lower curves (triangle markers) show the average ROI intensity in the directly registered volumes. The dark curve represents the true ROI intensity. The upper curves (star markers) represent the average ROI in registered volumes using intensity compensating high-pass filtering.

elastic motion in the 3D volumes by evaluating the displacement of each voxel as a Gaussian function of 4 independent parameters namely, 1. *time-point* (4,10), 2. *slice location* within the volume (8,10), 3. *rotation angle* (10°, 50) about the eye-center and 4. constant *translation* of 20 vertically. The values in the brackets specify the mean and standard deviation of the independent parameters. The rationale behind this procedure is that deformation is smooth and continuous spatially and relatively faster across the time-points. Further, the deformation in each time-point is maximum at a given slice and gradually declines for the neighboring slices. The translation may be due to blood flow or regular breathing and cannot be correlated across time-points and hence represented as a constant. The Gaussian parameters were chosen iteratively by a oncologist to reflect deformation found in practice. The deformed dataset is interpolated back onto a grid to represent a dataset with motion. The attached video sequences *static.avi* and *deform.avi* show the deformation obtained in a particular slice across all time-points. Using the original and deformed volumes, we can now evaluate the registration performance.

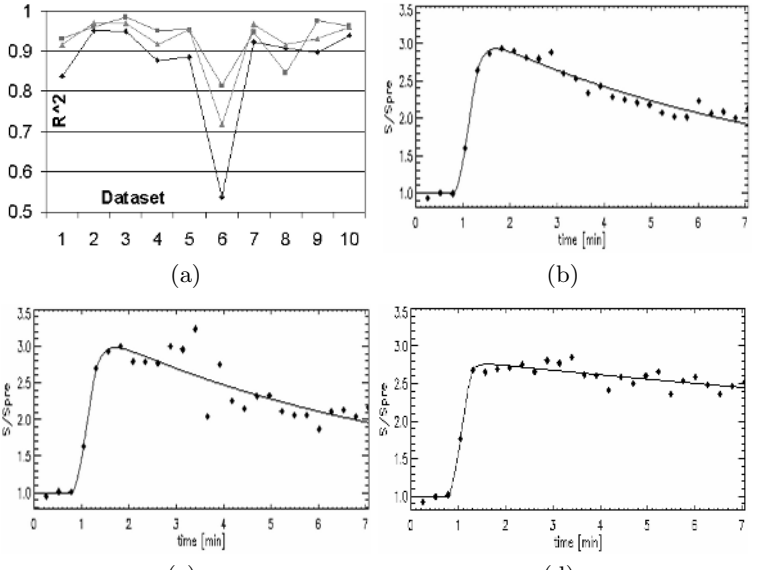
Figure 2(a) shows the Pearson's correlation plots of the eye dataset. The bottom (broken) curve indicates the correlation with the base volume (6^{th} epoch) prior to registration. The correlation in the volumes was 0.45 prior and 0.7 post-registration on average. Registration improves the correlation significantly. The low values are partly due to the motion, white noise in the images and also from the continuous uptake and washout of the contrast medium. The higher curves indicate the correlation values for 50, 100 and 200 iterations of the registration algorithm respectively. Longer iterative times tend to produce better registrations with diminishing returns progressively.

Figure 2(b) shows that as the registration accuracy improves (as a result of increased iteration), the drop in the average ROI intensity is more pronounced. The three lower curves (triangle markers) are directly registered volumes with-



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

Fig. 3. (a): Static slice. (b): Motion-affected slice. (c): Difference image. (d): Difference image after registration.



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

Fig. 4. (a): R^2 fit ratios using the pharmacokinetic model for the 3 cases in EYE-OS-P06 (8th slice). Fit observed using (b): multiple ROIs in the unregistered (c) single ROI in the unregistered and (d): single ROI in the new registered framework.

out any enhancements for 50, 100 and 200 iterations. In order to counter the smoothing during registration, we perform high pass filtering on the images prior to registration. However, this amplifies the shot noise to a great extent thereby defeating the purpose. Hence, the high-pass filter is preceded by a gradient-based anisotropic diffusion filter [15] that reduces the high frequency noise. Finally, we use a histogram matching procedure after the registration is complete to map the intensities back to the original image as shown in Figure 1. Here, the reference image is the image prior to registration. The histogram matching is done only in a region local to the ROI. We also threshold using mean intensity value in order to eliminate the background voxels from interfering. Using this new framework,

the registered volumes were scaled as shown by the curves with star markers in Figure 2. These curves now lie in the vicinity of the dark curve that represents the true ROI intensity and share the same trends.

A total of 4 iterations yielded satisfactory results with a time-step of 0.125 and the conductance term set to 1.0 for the gradient anisotropic filter. The high pass filter was implemented as a inverted 2D Gaussian filter with $\sigma = 0.75$. The normalized frequency domain representation is given by: $F(\omega_x, \omega_y) = N(1 - \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{{\omega_x}^2 + {\omega_y}^2}{\sigma^2}})$, where N is a normalizing factor to make the $Gain = 1$. Figure 3(a,b) shows two slices at the 6^{th} and 10^{th} time-points of the EYE-OS-P05 dataset. Slice 3(b) suffers from motion-induced warping. Figure 3(c) is the difference image of 3(a) and 3(b), showing the mis-alignment. The new difference image after registration is shown in Figure 3(d). The improvements and alignment of the boundaries are easy to observe over 3(c).

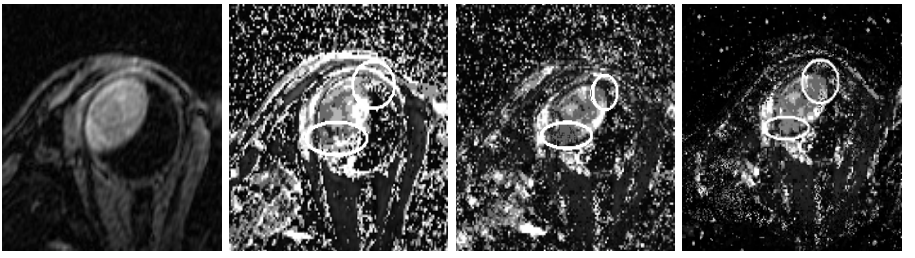
Why is filtering important? The crucial importance of DCE-MRI has been in understanding the heterogeneity in the tumor with respect to its vascular arrangement and density. Smoothing of images can potentially be dangerous since it causes the lesion to exhibit a homogenous texture. Preserving this important aspect requires intensity compensation by means of high-pass filtering and without amplifying the noise.

4 Experiments and Discussion

In this section, we strive to validate the performance of the registration in terms of the fit provided by the pharmacokinetic model. Three scenarios were considered: 1.) A single ROI was placed over the lesion for a given slice location where the tumor was well-defined and replicated at all other time points in the unregistered volumes. 2.) A trained pathologist placed ROIs for a given slice location consistently for all the time-points in the unregistered volumes. In other words, the ROI was shifted based on observed deformation. 3.) A single ROI was placed over the lesion at a given slice location and replicated at all other time points in the registered dataset. Here, there is no movement of the ROI.

The first case represents the assumption that there is no motion at all. The second situation realizes that there is motion and we manually correct the ROI placement. The third case assumes that there is motion and we use motion-correction procedures and then place a single ROI. We can therefore expect that the pharmacokinetic model would provide the best fit to the second case and the worst fit to the first case. We try to evaluate the fit that the third case provides.

Goodness of fit Statistics: The fit provided by the model to the experimental data obtained from the three cases was evaluated using the parameter $R^2 = 1 - \frac{SS_{error}}{SS_{mean}}$, where SS_{error} is the sum of the squares of the distance between the experimental data and the best-fit curve and SS_{mean} is the sum of squares from the mean of all points [16]. When $R = 0$, the fit provided is no better than the mean horizontal line. When $R = 1$, the curve fits the points exactly.



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

Fig. 5. (a): EYE-OS-P25 with tumor (b): Unregistered (c): Registered dataset without enhancement (d) Registered dataset with enhancement

Analysis: We picked 10 datasets randomly and plot the R^2 fit obtained for the 3 cases in Figure 4 (a). As expected, the fit obtained from case 2 (square markers) is better than case 1 (diamond markers). Case 3 (triangle markers) performed better than the case 1 and matched or outperformed case 2 (the ideal case). Hence, the registered dataset with a single ROI consistently encompassed the tumor region for the given pharmacokinetic model. An example of the fits for the EYE-OS-P06 is plotted for illustration. The model fits (Figure 4d - case 3; 4b - case 2) better than (4c - case 1).

Pixel-mapped model: While the above experiments were based on average signal intensities drawn from an ROI, we now investigate the results on applying the model to each voxel in the $3D$ volume. The two important lesion parameters (A and k_{ep}) are intensity-mapped and plotted. The noisy pixel-wise analysis shows that the lesion has intense local enhancement with spot noise. Figure 5 shows the pixel mappings for the tumor in EYE-OS-P25 dataset. We make three observations: 1) In the unregistered dataset, the tissue structures such as the upper rim of the eye are more well-defined in terms of the enhancement locally. The structures are however absent in the directly registered case. The reason for this is the smoothing that happens in registration destroys the thin structures and reduces the enhancement pattern per pixel. 2) The registered datasets (c,d) define the tumor boundaries better than the unregistered case especially in the upper and lower poles (white circles). It indicates that there are more pixels that are enhanced indicating a better fit to the model. 3) The registered dataset with the high pass filtering improves over the directly registered case by reducing the shot noise, mapping the other peripheral eye structures and providing a better characterization of the tumor interior.

In conclusion, the effect of blurring in the pixel intensities and on the lesion parameters is explored. We incorporate filtering operations prior to registration to reduce noise and pre-empt signal degradation. Our validation covers stationary datasets with induced motion for ground-truth. The framework is tested on clinical cases. We found that motion artificially changes lesion morphology near the boundaries. Using our new framework, the effect for contrasting differences in the enhancement pattern for quantitative evaluation was pronounced.

References

1. Taylor, J.S., Tofts, P.S., Port, R., et al: MR imaging of tumor microcirculation: Promise for the new millennium. J. Magn. Reson. Imaging **10** (1999) 903–907
2. Knopp, M.V., Weiss, E., Sinn, H.P., et al: Pathophysiologic basis of contrast enhancement in breast tumors. J. Magn. Reso. Imaging **10** (1999) 260–266
3. Hayton, P., Brady, M., Tarassenkoa, L., et al: Analysis of dynamic MR breast images using a model of contrast enhancement. Med. Image Anal. **1** (1997) 207–224
4. Bidaut, L.M., Vallee, J.P.: Automated registration of dynamic MR images for the quantification of myocardial perfusion. J. Magn. Reson. Imaging **13** (2001) 648–655
5. Buonaccorsi, G.A., Roberts, C., Cheung, S., et al: Tracer kinetic model-driven registration for dynamic contrast enhanced MRI time series. In: MICCAI'05, Lect. Notes in Comp. Sc. (2005)
6. Lucht, R., Knopp, M.V., Brix, G.: Elastic matching of dynamic MR mammographic images. Magn. Reson. in Medicine **43** (2000) 9–16
7. Rueckert, D., Sonoda, L., Hayes, C., et al: Nonrigid registration using free-form deformations: application to breast MR images. IEEE Trans. Med. Imag. **19** (1999) 712–721
8. Maes, F., Vandermeulen, D., Suetens, P.: Medical image registration using mutual information. Proceedings of the IEEE **91(10)** (1996) 1699–1722
9. Roche, A., Malandain, G., Pennec, X., et al: The correlation ratio as a new similarity measure for multimodal image registration. Lect. Notes Comput. Sc. **1496** (1998) 1115–1124
10. Thirion, J.P., Calmon, G.: Measuring lesion growth from 3D medical images. In: Nonrigid and Articulated Motion Workshop (NAM'97). (1997)
11. Thirion, J.P.: Fast non-rigid matching of 3D medical images. In: Medical Robotics and computer aided surgery, MRCAS'95. (1995) 47–54
12. Thirion, J.P.: Fast non-rigid matching of 3D medical images. In: Computer Vision and Pattern Recognition, CVPR'96. (1996)
13. Thirion, J.P., Subsol, G., Dean, D.: Cross validation of three inter-patients matching methods. In: Visualization in Biomedical Computing VBC'96, Lect. Notes in Comp. Sc. Volume 1131. (1996) 327–336
14. Ibáñez, L., Schroeder, W.: The ITK Software Guide. The Insight and Registration Toolkit [www.itk.org], Kitware, Inc. (2003)
15. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Trans. on Pattern Anal. and Mach. Intell. **12** (1990) 629–639
16. Fan, X., Medved, M., River, J.N., et al: New model for analysis of dynamic contrast-enhanced mri data distinguishes metastatic from nonmetastatic transplanted rodent prostrate tumors. Magn. Reson. in Med. **51** (2004) 487–494

Automated Analysis of the Mitotic Phases of Human Cells in 3D Fluorescence Microscopy Image Sequences

Nathalie Harder[1], Felipe Mora-Bermúdez[2], William J. Godinez[1], Jan Ellenberg[2], Roland Eils[1], and Karl Rohr[1]

[1] University of Heidelberg, IPMB, and DKFZ Heidelberg, Dept. Bioinformatics and Functional Genomics, Im Neuenheimer Feld 364, D-69120 Heidelberg, Germany
[2] European Molecular Biology Laboratory (EMBL), Gene Expression and Cell Biology/Biophysics Programmes, Meyerhofstrasse 1, D-69117 Heidelberg, Germany
n.harder@dkfz-heidelberg.de

Abstract. The evaluation of fluorescence microscopy images acquired in high-throughput cell phenotype screens constitutes a substantial bottleneck and motivates the development of automated image analysis methods. Here we introduce a computational scheme to process 3D multicell time-lapse images as they are produced in large-scale RNAi experiments. We describe an approach to automatically segment, track, and classify cell nuclei into different mitotic phases. This enables automated analysis of the duration of single phases of the cell life cycle and thus the identification of cell cultures that show an abnormal mitotic behavior. Our scheme proves a high accuracy, suggesting a promising future for automating the evaluation of high-throughput experiments.

1 Introduction

In recent years the technology of *RNA interference* (RNAi) has become the method of choice for identifying the biological function of genes in the field of functional genomics. With this screening method all known genes of an organism are systematically silenced one after the other, and the resulting morphological changes are analyzed. However, such large-scale knockdown screens provide enormous amounts of data which require tools for automated image analysis.

Our work is carried out within the EU project MitoCheck, which aims to explore the coordination of mitotic processes in human cells at a molecular level and to contribute to revealing the mechanisms of cancer development. To identify the genes that are involved in cell division (mitosis), genome-wide high-throughput RNAi primary screens are performed. In addition, RNAi secondary screens are performed which have a higher spatial and temporal resolution and involve several hundred genes. Fluorescence microscopy time-lapse images of the treated cell culture are acquired to study the effect of a silenced gene on mitosis. This contribution is concerned with the automated evaluation of an assay to analyze the duration of the different phases in mitotic progression. In *normal* cells of one type the duration of the mitotic phases is similar. To study the duration of mitotic phases of *treated* cells a specific assay has been developed which is

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 840–848, 2006.

characterized by significant elongations of mitotic phases compared to control experiments. To automatically determine whether mitotic phases of treated cells are longer compared to normal cells, cells have to be observed throughout their life cycle and for each time point the respective phase has to be determined.

Previous work on automated analysis of cell images has been done in different application fields. Based on fluorescence microscopy imaging complete cells as well as single subcellular structures have been studied. The automated recognition of subcellular structures is a major task in location proteomics and work has been done in this field given 2D (e.g., [1, 2]) and 3D (e.g., [3]) single cell images. Classification of complete cells has been performed, for example, to investigate the influence of drugs on cellular proteins [4]. In [5] tracking of cells in 2D microscopy images has been used to improve segmentation and cell quantification. Automated analysis of cell images plays an increasing role particularly for the evaluation of high-throughput cell phenotype screens. Approaches for *single-frame* multi-cell *2D* images have been described in [6, 7, 8].

We have developed an approach to analyze multi-cell *image sequences* from large-scale RNAi secondary screens. In comparison to previous work, we analyze *3D* cell array time-lapse images that include multiple cell nuclei in *different mitotic phases*. Each image of an image sequence contains three slices that have been acquired with a confocal fluorescence microscope. Our computational scheme for analyzing the duration of mitotic phases consists of four main steps: Segmentation of multi-cell images, tracking of cells, image feature extraction, and classification into the mitotic phases. For fast and accurate segmentation we use a region adaptive thresholding technique. Our tracking scheme is able to cope with the splitting of cells during mitosis which is important in our application. Based on the tracking result we automatically select the most informative slice out of the 3D image, which is then used for feature extraction. Compared to previous work, we incorporate temporal changes of the cell morphology between ancestrally related cells by directly including dynamic image features. A Support Vector Machine classifier is used to classify the cells into the following seven mitotic phases: *Interphase, Prophase, Prometaphase, Metaphase, Anaphase1, Anaphase2*, and *Telophase*. Based on the classification result in subsequent images the duration of the different phases can be determined. It is the first time that an automated scheme is presented which yields a detailed classification into the above mentioned phases of dividing cells. Our approach has been successfully applied to four image sequences from large-scale RNAi screens. We have compared the performance with ground truth provided by manual evaluation and it turned out that we obtain an average accuracy of around 89% and an overall accuracy of 94.6% for the classification.

2 Methods

2.1 Image Analysis Workflow

For a detailed analysis of the different phases of dividing cells high-resolution confocal fluorescence microscopy images of the DNA have been acquired. Compared

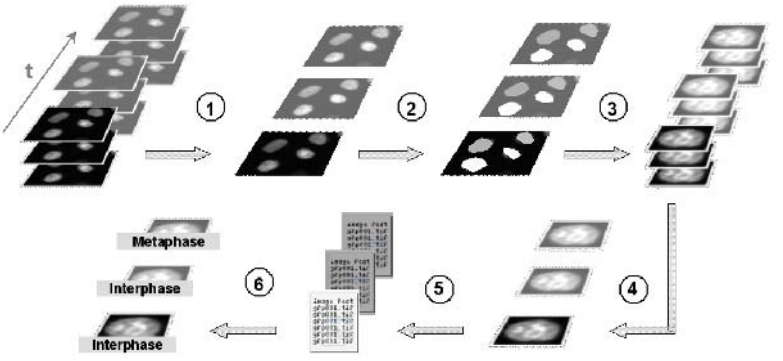


Fig. 1. Image analysis workflow: (1) Maximum intensity projection of multi-cell 3D images, (2) Segmentation and tracking in 2D, (3) Extraction of 3D ROIs including single cells, (4) Selection of most informative slice, (5) Feature extraction, (6) Classification

to previous work we here use 3D images consisting of three confocal planes (slices). The reason to use several confocal planes is that cells change their morphology during mitosis, i.e., in the interphase they are flat, but with the beginning of cell division they arch upward and take the shape of a hemisphere. Therefore, if we would use the same focal plane throughout the cell cycle the cell structure (DNA) could not well be observed. Since we are using multi-cell images that contain cells in different mitotic phases, it is impossible to define one focal plane per time step that well represents all cells. Therefore, three slices from three different focal planes are acquired that cover the range of possible cell shape changes. Because of technical reasons in the image acquisition process the number of slices is restricted to three. This relatively small number of slices does not allow a complete 3D analysis. Thus, we apply the workflow shown in Fig. 1. In the first step, we apply a maximum intensity projection (MIP) for each time step resulting in 2D images. In the second step, we perform segmentation and tracking based on these MIP images to determine the correspondences of cell nuclei in subsequent frames. Based on the segmentation and tracking result we now go back to the original 3D images and define 3D ROIs for each cell. For each 3D ROI we choose the slice containing the most information. The reason of using the original images lies in the fact that these images contain more detailed information about the inner intensity structure of a cell than the MIP images. Finally, static and dynamic features are extracted and classification is performed resulting in a sequence of mitotic phases for each cell trajectory.

2.2 Segmentation of Multi-cell Images

Since we have to cope with a huge number of multi-cell images, fast and reliable segmentation of single objects is crucial. Various advanced segmentation algorithms have been described in the literature, but as computation time plays an important role in our application, the speed of the algorithm is a decisive criterion. We have investigated three different thresholding techniques which we

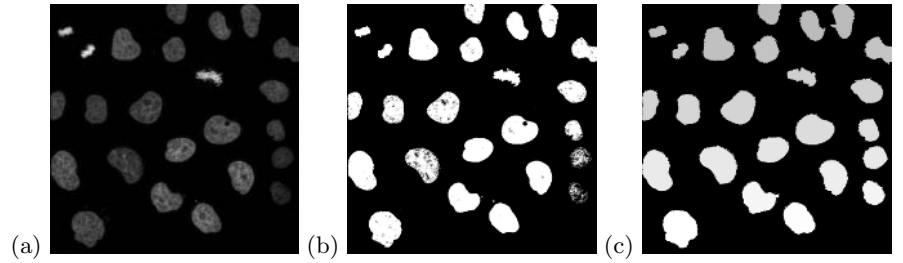


Fig. 2. (a) Original image, (b) Global threshold (Otsu), (c) Local adaptive threshold

have previously used for analyzing static 2D multi-cell images. It turned out that global thresholding alone does not produce satisfying results especially for low-contrast cells. Instead, local adaptive thresholding can segment the cell nuclei with high accuracy (see Fig. 2). Consequently, we apply the latter technique which incorporates several improvements compared to [8]. The algorithm uses a quadratic sliding window to calculate local thresholds for different image regions. A local threshold is only calculated if the variance within the window reaches a user-defined threshold, else a global threshold is used [9]. This ensures that only for regions that contain some information the local threshold is calculated, which significantly reduces the computation time. The threshold is applied to a quadratic region at the center of the sliding window and the window is shifted by the length of this inner region. Compared to [8] this modification reduces segmentation errors at the region borders. To calculate the global and the local intensity thresholds we applied two histogram-based threshold selection schemes: *Otsu* thresholding [10] and *minimum error* thresholding [11]. We found that the combination of [10] to determine the global threshold and [11] to determine the local thresholds yielded the best results (see Fig. 2c).

2.3 Tracking of Mitotic Cell Nuclei

To analyze the mitotic behavior of single cells in multi-cell images, a tracking scheme is required that determines the temporal connections and can handle splitting objects. We have developed the following two-step tracking scheme: First, initial, non-splitting trajectories are established, and second, mitotic events are detected and the related trajectories are merged. In the first step, the initial trajectories are determined using a feature point tracking algorithm based on [12]. As feature points we use the centers of gravity of segmented cell nuclei. For each frame of an image sequence the algorithm considers the predecessor and successor frame. In these frames, object correspondences are determined by searching for trajectories with maximum smoothness. In the second step, all trajectories that do not start in the first frame are taken into account as possible mitotic events. The decision whether a trajectory should be treated as mitosis event is made depending on the distance-overlap-ratio of potential parent and child objects [13]. In addition, the smoothness of the potential compound trajectory is taken into account. For the identified splitting events the corresponding tracks are finally merged. As an example, Fig. 3 shows a tracking result.

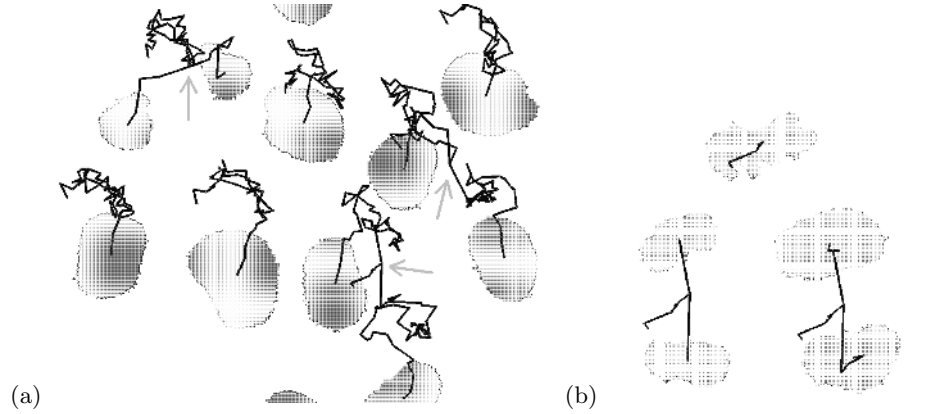


Fig. 3. (a) Tracking result including three mitotic cells, (b) Example of the splitting of one cell in three consecutive images (top, bottom left, bottom right)

2.4 Extraction of Static and Dynamic Features

As described in Section 2.1, we determine for each cell nucleus 3D ROIs in the original 3D images. Within these ROIs we select the cell's individual *most informative slice* in order to compute image features. Two selection criteria for determining the optimal slice have been tested: Maximum total intensity and maximum entropy. We found that the *maximum-intensity* criterion performs very well compared to manual selection, whereas the *maximum-entropy* criterion often fails and shows a high sensitivity to noise. Consequently, we here apply the *maximum-intensity* criterion. Within the selected slice we compute a set of static and dynamic image features. The static features comprise granularity features, object- and edge-related features, tree-structured wavelet features, Haralick texture features, grey scale invariants, and Zernike moments. The dynamic features represent the morphological changes of cells in adjacent frames. To this end, we compute the difference of three basic image features (*object size, mean intensity,* and *standard deviation of the intensity*) for each cell to its predecessor and to its successor. We chose these three features because they exhibit characteristic changes during mitosis. As an example, Fig. 4 shows a plot of the mean intensity over time for a dividing cell. It can be seen that, for example, between *Prophase* and *Prometaphase* there is a steep rise of the mean intensity and for *Anaphase1* there is a characteristic spike. In total, we compute 332 features for each cell nucleus. For the training set, we standardize each feature w.r.t. a mean value of zero and a standard deviation of one. In the test set, the feature values are linearly transformed based on the transformation parameters of the training set.

2.5 Classification

We apply a Support Vector Machine (SVM) classifier with a Radial Basis Function (RBF) as kernel function to classify the nuclei into the seven classes *Interphase, Prophase, Prometaphase, Metaphase, Anaphase1, Anaphase2,* and

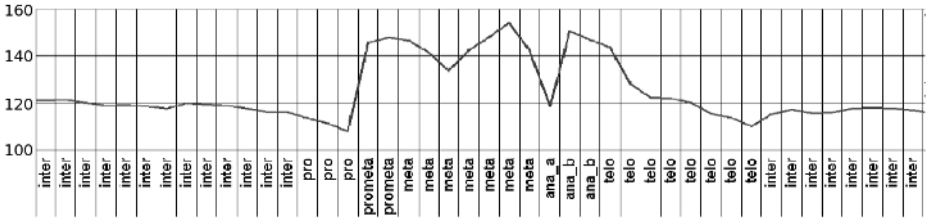


Fig. 4. Mean intensity for a dividing cell and manually assigned phases of mitosis

Telo-phase (see Fig. 5). SVMs are mathematically well-founded and they have the advantage that their complexity is independent of the dimension in feature space compared to other classification methods. This property allows us to work with a high number of features and we can skip the feature selection step as it is not crucial. We solve the multi-class classification problem with a "one-against-one" approach. To optimize the penalty parameter C and the kernel parameter γ for the Radial Basis Function, we perform a three-fold cross-validation with varying values of C and γ on the training set (model selection) prior to the actual training of the classifier.

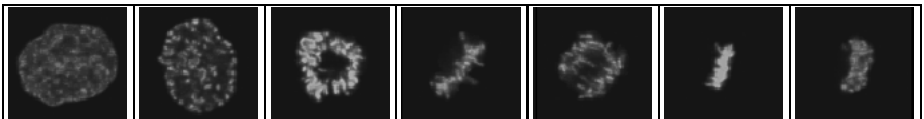
3 Experimental Results

3.1 Image Data

In our study we use 3D image sequences that have been acquired with a confocal fluorescence microscope in the framework of the EU project MitoCheck at the EMBL Heidelberg. Chromosome morphology is visualized using a HeLa (Kyoto) cell line stably expressing the chromosomal marker histone 2B-EGFP. One image contains about 20 nuclei with an average diameter of approximately 100-150 pixels (in *Interphase*). All images have a gray value depth of 8 bit and a spatial resolution of 1024×1024 pixels. Each image sequence comprises 124 to 128 time steps with a temporal resolution of 7 min and three image slices per time step.

3.2 Classification Results

We have applied our approach given four 3D multi-cell image sequences. Using maximum intensity projection for all stacks of all sequences resulted in 500



Interph.	Proph.	Prometaph.	Metaph.	Anaph.1	Anaph.2	Teloph.
1143	*79*	*56*	*120*	*17*	*89*	*263*

Fig. 5. Example images of the different mitotic phases and number of available samples

Table 1. Confusion matrix and accuracies for the experiment with five-fold cross validation; dynamic features have *not been included*; overall accuracy: 92.9%

True Class	**Classifier Output**							**Accur.**
	Inter	Pro	Prometa	Meta	Ana1	Ana2	Telo	[%]
Inter	**1109**	8	0	1	0	0	26	97.0
Pro	11	**66**	1	0	1	0	0	83.5
Prometa	0	0	**44**	6	6	0	0	78.6
Meta	1	0	5	**110**	0	0	4	91.7
Ana1	1	0	5	3	**5**	2	1	29.4
Ana2	0	0	0	0	1	**76**	12	85.4
Telo	21	0	0	1	2	8	**231**	87.8

Table 2. Same as Tab. 1 but *including* dynamic features; overall accuracy: 94.6%

True Class	**Classifier Output**							**Accur.**
	Inter	Pro	Prometa	Meta	Ana1	Ana2	Telo	[%]
Inter	**1114**	6	0	1	0	0	22	97.5
Pro	5	**72**	2	0	0	0	0	91.1
Prometa	0	0	**51**	5	0	0	0	91.1
Meta	0	0	2	**114**	1	1	2	95.0
Ana1	1	0	2	2	**11**	1	0	64.7
Ana2	0	0	0	0	1	**84**	4	94.4
Telo	33	0	0	1	1	3	**225**	85.6

projected multi-cell images. These images have been segmented and tracked. The tracking scheme was able to detect 80.0% of all occurring mitosis events (determined by visual inspection). Since the subsequent processing steps rely on correctly detected mitosis events we corrected the remaining ones manually. Static and dynamic features have been extracted within the most informative slices for all segmented and tracked nuclei. To provide ground truth, all nuclei have been classified manually into the seven phases of mitosis. We split the available samples for each class randomly into training data and test data at a ratio of 2:1. The numbers of samples and example images for each class are given in Fig. 5. The feature values were standardized and a Support Vector Machine classifier was trained as described above. For the different phases (classes) we obtain between 80.0% and 98.1% correct classifications and an average accuracy of 90.1% (data not shown). To check the reliability of the result we repeated the classification step, applying a five-fold outer cross-validation on the whole data set. As shown in Tab. 2, we obtain classification accuracies of 64.7% to 97.5% for the different phases and an average accuracy of 88.5%. Thus, both average accuracies correspond well and we can draw the conclusion that we can rely on an average classification accuracy of around 89%. If we do not just average the accuracies for the different classes but take into account the number of samples per class we obtain an overall accuracy of 94.6%.

To examine the effect of including dynamic features, we removed all dynamic features and repeated the classification experiments. In this case, we obtained an average accuracy of 79.1% and an overall accuracy of 92.9% for the experiment

with five-fold cross-validation (see Tab. 1). It turns out that by *including* dynamic features we can improve the result significantly. The improvement becomes even more obvious when we compare the accuracies for the single phases of both experiments as listed in Tabs. 1 and 2: If dynamic features are included the accuracies are significantly higher for all phases, except for *Telophase*. The largest improvement can be observed for *Anaphase1* with an increase from 29.4% to 64.7% and for *Prometaphase* with an increase from 78.6% to 91.1%.

4 Discussion and Conclusion

We have presented an approach for automated analysis of the duration of mitotic phases in 3D confocal microscopy image sequences. Our approach segments and tracks splitting cells throughout the sequences and thus determines tree-structured tracks which represent cell pedigrees. By using static and dynamic features our scheme classifies the cells with high accuracy into seven mitotic phases. Based on the sequence of phases that have been computed in subsequent frames, the duration of each phase can be determined. The computed phase durations can be compared to control experiments to automatically identify experiments that show delays in mitotic phases.

Acknowledgement

This work has been supported by the EU project MitoCheck.

References

1. Huang, K., Murphy, R.: Boosting accuracy of automated classification of fluorescence microscope images for location proteomics. BMC Bioinf. **5** (2004) 78
2. Conrad, C., Erfle, H., Warnat, P., Daigle, N., Löch, T., Ellenberg, J., Pepperkok, R., Eils, R.: Automatic identification of subcellular phenotypes on human cell arrays. Genome Research **14** (2004) 1130–1136
3. Chen, X., Murphy, R.: Robust classification of subcellular location patterns in high resolution 3D fluorescence microscope images. In: Proc. 26th Int. Conf. IEEE Engineering in Medicine and Biology Society, San Francisco, CA (2004) 1632–1635
4. Lindblad, J., Wählby, C., Bengtsson, E., Zaltsman, A.: Image analysis for automatic segmentation of cytoplasms and classification of Rac1 activation. Cytometry Part A **57A** (2003) 22–33
5. Yang, F., Mackey, M., Ianzini, F., Gallardo, G., Sonka, M.: Cell segmentation, tracking, and mitosis detection using temporal context. In: Proc. MICCAI'2005. Volume 3749., Palm Springs, CA, USA, Springer-Verlag (2005) 302–309
6. Perlman, Z., Slack, M., Feng, Y., Mitchison, T., Wu, L., Altschuler, S.: Multidimensional drug profiling by automated microscopy. Science **306** (2004) 1194–1198
7. Zhou, X., Liu, K.Y., Bradley, P., Perrimon, N., Wong, S.T.: Towards automated cellular image segmentation for RNAi genome-wide screening. In: Proc. MICCAI'2005. Volume 3749., Palm Springs, CA, USA, Springer-Verlag (2005) 302–309

8. Harder, N., Neumann, B., Held, M., Liebel, U., Erfle, H., Ellenberg, J., Eils, R., Rohr, K.: Automated recognition of mitotic patterns in fluorescence microscopy images of human cells. In: Proc. ISBI'2006, Arlington, USA (2006) 1016–1019
9. Gonzalez, R., Woods, R.: Digital Image Processing. Prentice Hall (2002)
10. Otsu, N.: A threshold selection method from grey level histograms. IEEE Transactions on Systems, Man and Cybernetics **9** (1979) 62–66
11. Kittler, J., Illingworth, J.: Minimum error thresholding. Pattern Recognition **19** (1986) 41–47
12. Chetverikov, D., Verestoy, J.: Tracking feature points: a new algorithm. In: Proc. 14th Int. Conf. Pattern Recognition, Brisbane, Australia (1998) 1436–1438
13. Withers, J.: Tracking cell splits and merges. In: Southwest Symp. Image Analysis and Interpretation. (1996) 117–122

Spline-Based Probabilistic Model for Anatomical Landmark Detection

Camille Izard[1,2,⋆], Bruno Jedynak[1,2], and Craig E. L. Stark[3]

[1] Laboratoire Paul Painlevé, Université des Sciences et Technologies de Lille, France
[2] Center for Imaging Science, Johns Hopkins University, Baltimore, MD
[3] Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD
{camille.izard, bruno.jedynak, ctark}@jhu.edu

Abstract. In medical imaging, finding landmarks that provide biologically meaningful correspondences is often a challenging and time-consuming manual task. In this paper we propose a generic and simple algorithm for landmarking non-cortical brain structures automatically. We use a probabilistic model of the image intensities based on the deformation of a tissue probability map, learned from a training set of hand-landmarked images. In this setting, estimating the location of the landmarks in a new image is equivalent to finding, by likelihood maximization, the "best" deformation from the tissue probability map to the image. The resulting algorithm is able to handle arbitrary types and numbers of landmarks. We demonstrate our algorithm on the detection of 3 landmarks of the hippocampus in brain MR images.

1 Introduction

Anatomical landmarks are well-defined points in the anatomy that experts use to establish biologically meaningful correspondences between structures [1]. Such correspondences are commonly used by registration algorithms, as initialization and/or as constraints [2, 3, 4]. Landmarks also provide a local shape description useful for anatomical shape comparison [5].

However, locating landmarks on biological structures is a challenging and time-consuming task, even for experts. This has motivated the development of several methods for automatic landmarking. Previously proposed methods use either 3D filters to detect, for example, high curvature points and corners in the image [6], or a geometric model of the image intensities to detect, for example, the tip of a structure [7, 8]. Both techniques rely on local intensity variations, but some landmarks are not detectable using only intensity information. This is the case of the head of the hippocampus whose intensity is similar to the surrounding amygdala. Notice also that these techniques are designed for the detection of independent landmarks. They cannot integrate information from the previous detections, although it should be helpful to know, for example, the location of the head of the hippocampus to locate its tail.

⋆ Supported by the Doctoral Fellowship of the Université des Sciences et Technologies de Lille, by ARO/DAAD19/-02-1-0337, by NIH, ADRC Pilot Project Award, 2005 and by general funds from the Center for Imaging Science of Johns Hopkins University.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 849–856, 2006.

In this paper, we propose a probabilistic approach[1] to landmark brain structures such as the hippocampus, on which one needs to locate the head, the tail and the extremity of the hippocampal uncus. We identify the location of the landmarks in an image with a unique deformation of the underlying 3D space. The set of deformations considered in this paper is a set of Gaussian interpolating splines where the landmarks act as control points. We then build a probabilistic model for an image, given the landmark locations. The estimation of the model parameters consists of learning a local tissue probability map, using a training set of hand-landmarked images. In new images, landmarks are identified using a gradient ascent algorithm on the likelihood function.

The resulting algorithm is generic and specializes automatically to the structure or region of interest during the learning of the tissue probability map. Therefore the algorithm is able to specialize to arbitrary types and numbers of landmarks.

In section 2, we describe the generative model of the image intensities. In section 3, we show how to estimate the parameters of this model, including the photometric parameters of the tissue types and the local tissue probability map. In section 4, we show how the likelihood function can be maximized to estimate the position of the landmarks in a new image. Finally, in section 5, we test the method on the simultaneous detection of 3 landmarks of the hippocampus.

2 Generative Model of the Image Intensities

In this section, we describe the generative model of the intensities of image i, $\boldsymbol{X}^{(i)}$. Let $\boldsymbol{L}^* = (L_1^*, \cdots, L_K^*)$ and $\boldsymbol{L} = (L_1, \cdots, L_K)$, be two sets of K landmarks in $\mathbb{R}^3$ and $\phi : \mathbb{R}^3 \to \mathbb{R}^3$ a small deformation such that $\phi(\boldsymbol{L}^*) = \boldsymbol{L}$. Since many deformations verify this condition, we restrict the set of deformations to a Gaussian spline interpolation of the landmarks displacements, see equation (3). We fix a standard configuration $\boldsymbol{L}^*$, the center of mass of landmark locations in the training set. Hence finding the position of the landmarks $\boldsymbol{L}$ is equivalent to estimating the "best" deformation from $\boldsymbol{L}^*$ to $\boldsymbol{L}$. By the Bayes' formula the joint distribution becomes $P(\boldsymbol{X}^{(i)}, \phi) = P(\boldsymbol{X}^{(i)}|\phi)P(\phi)$. We choose a uniform prior so that maximizing the joint distribution is equivalent to maximizing the conditional distribution $P(\boldsymbol{X}^{(i)}|\phi)$.

We make the simplifying assumption that the voxel intensities are independent given the transformation ϕ. That is, $P(\boldsymbol{X}^{(i)}|\phi) = \prod_{v \in V} P(X_v^{(i)} = x|\phi)$, with $X_v^{(i)}$ the intensity at voxel v in image i and V the set of voxels in the image. The image intensities are modeled with a mixture of 6 gaussian distributions, corresponding to the following tissues: CSF, CSF+GM, GM, GM+WM, WM and skull+blood vessels. We denote by Z_v the discrete random variable representing the tissue type at voxel v. We assume that the intensity $X_v^{(i)}$, given the tissue type Z_v is independent of the deformation and write the conditional probability of the ith image given the deformation as

$$P(\boldsymbol{X}^{(i)}|\phi) = \prod_{v \in V} \sum_{j=1}^{6} P(X_v^{(i)} = x|Z_v = j)P(Z_v = j|\phi). \quad (1)$$

[1] In [9, 10], probabilistic models of the image intensities have also been proposed to segment and register brain MRI.

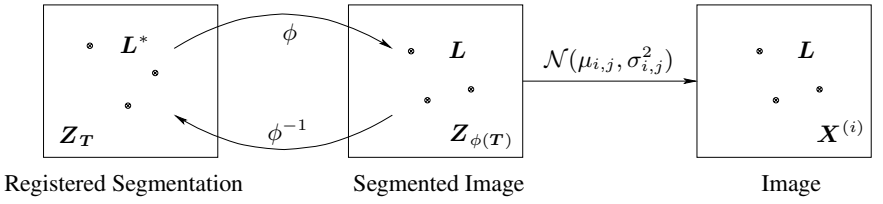


Fig. 1. To generate a new image $\boldsymbol{X}^{(i)}$ (rightmost): draw a random segmentation (leftmost) based on the distribution of the tissue probability map; apply a random deformation ϕ to find the new image segmentation $\boldsymbol{Z}_{\phi(\boldsymbol{T})}$ (middle) and assign an intensity chosen with the corresponding Gaussian distribution

The first term characterizes the photometry of each tissue in the image, while the second term encodes the geometry. It gives the probability of observing each tissue type at each location in the brain. Figure 1 illustrate the generative model.

Let β be the parameters of the transformation ϕ, the estimation of the landmark position in a new image is $\hat{\boldsymbol{L}} = \phi_{\hat{\beta}}(\boldsymbol{L}^*)$ with $\hat{\beta} = \arg\max_\beta \ln P(\boldsymbol{X}|\phi_\beta)$. We denote by $\boldsymbol{T} = \{P(Z_t = j), 1 \leq j \leq 6\}_{\forall t \in \boldsymbol{T}}$ the tissue probability map, i.e. the probability of observing tissue j at each voxel. Algorithm 1 summarizes the algorithm for automatic landmarking which will be described in detail in the following sections.

Algorithm 1 . Automatic Landmarking

Learning step

Given N training images with K landmarks $\boldsymbol{L}_i = (L_{i,1}, \cdots, L_{i,K}), 1 \leq i \leq N$:

1. *Identify the standard configuration:* $\boldsymbol{L}^* = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{L}_i$,
2. *Learn the photometry parameters:* $\mu_{i,j}$ and $\sigma^2_{i,j}$ for each tissue type j and image i,
3. *Register the images of the training set:*
   (a) for each image i, find the unique ϕ_{β_i}, such that $\phi_{\beta_i}(\boldsymbol{L}^*) = \boldsymbol{L}_i$,
   (b) for each t in the probability map, extract the intensities $x^{(i)}_{\phi_{\beta_i}(t)}$ for all i, $1 \leq i \leq N$,
4. *Estimate the tissue probability map:*
   for each t in the probability map, estimate the proportions of each tissue in $(x^{(1)}_{\phi_{\beta_1}(t)}, \cdots, x^{(N)}_{\phi_{\beta_1}(t)})$, using the EM algorithm.

For a new image

1. *Learn the photometry parameters:* μ_j and σ^2_j for each tissue type j,
2. *Estimate the transformation and predict the landmark location:*
   (a) Maximize the likelihood with respect to β using a gradient method,
   (b) Compute the transformation $\phi_{\hat{\beta}}$ and the predicted landmark location $\hat{\boldsymbol{L}} = \phi_{\hat{\beta}}(\boldsymbol{L}^*)$.

3 Estimation of the Model Parameters

In this section, we show how to estimate the parameters of the probabilistic model. Since the photometry and the geometry are assumed to be independent in our model,

we can estimate the photometric parameters and the tissue probability map separately. Let us start with the photometry.

3.1 Learning the Photometric Parameters

The intensity of an image $\boldsymbol{X}^{(i)}$ is modeled as a mixture of Gaussian distributions, assuming independence of the voxels, as it is commonly done in literature [11]. Thus,

$$P(\boldsymbol{X}^{(i)}) = \prod_{v \in V} \sum_{j=1}^{6} g_{i,j}(x_v^{(i)}) \alpha_{i,j}, \text{ with } g_{i,j} \sim \mathcal{N}(\mu_{i,j}, \sigma_{i,j}) \text{ and } \sum_j \alpha_{i,j} = 1. \quad (2)$$

Given the segmentation, it would be straightforward to estimate the parameters of the model. However, here the segmentation is unknown, also we use the Expectation-Maximization (EM) algorithm [12] to maximize $\ln P(\boldsymbol{X}^{(i)})$ with respect to $\mu_{i,j}$, $\sigma_{i,j}^2$, and $\alpha_{i,j}$, $1 \leq j \leq 6$. In the case of a Gaussian mixture, both the E-step and the M-step can be written in closed form and convergence to a local maximum of the likelihood function is guaranteed.

3.2 Learning the Tissue Probability Map

The standard atlas for registration in neuroimaging is that of Talairach and Tournoux, [2]. Tissue probability maps have been estimated based on images aligned by linear transformation to that atlas [13]. In the problem we consider the images are already manually transformed into Talairach space but we try to reduce the residual variability (we note variability of 10mm for the head of the hippocampus [14]). Also we use the correspondences of the landmarks of the training set to improve the quality of the alignment of the structure of interest and consequently learn a locally more precise probability map. The estimation of the tissue probability map is obtained as follows: first, register the training images so that the landmarks lie at the same location in all the images and then estimate the proportions of each tissue type at each location.

Registration of the Training Set. Since the images have been previously aligned by a linear transformation, we deal only with small deformations. More specifically we use a spline interpolation ϕ_β based on the landmark matching constraint.

$$\forall t \in \mathbb{R}^3, \phi_\beta(t) = t + \sum_{k=1}^{K} \frac{\beta_k}{\sqrt{2\pi}^3 \sigma_k^3} \exp\left(-\frac{\|t - L_k^*\|^2}{2\sigma_k^2}\right), \quad (3)$$

where $\beta_k \in \mathbb{R}^3$ is the displacement vector at the landmark L_k^* and $\sigma_k \in \mathbb{R}$ is the smoothing parameter. This type of deformation has a simple analytical form, easy to generalize to more landmarks. It also has a simple interpretation and an almost local support so that computation will be easier. Registering the training images consists of finding the unique transformation ϕ_β that transforms the landmarks of the tissue probability map to those of the image, i.e. $\phi_\beta(\boldsymbol{L}^*) = \boldsymbol{L}$. The choice of the smoothing parameters σ_k is done manually, so that we ensure the invertibility of the deformation. These parameters could also be learned during the training phase. Once σ_k are fixed,

registering one training image is equivalent to solving for β the linear system given by the landmark matching constraints. Other deformations could be used to register the images which in addition to satisfying the landmark matching constraints, either minimize the bending energy (Thin-Plate Splines) or ensure invertibility (Geodesic Interpolating Splines) [15], at some computational cost.

Estimation of the Probability Tissue Map. The registration of the training set provides us a vector of intensities at each location $t \in \boldsymbol{T}$, $\left(x^{(1)}_{\phi_{\beta_1}(t)}, \ldots, x^{(N)}_{\phi_{\beta_N}(t)}\right)$, where $x^{(i)}_{\phi_{\beta_i}(t)}$ is the gray level of the image i, $1 \leq i \leq N$, at the location $\phi_{\beta_i}(t)$. Learning the geometry consists of estimating the proportion of each tissue type j at each location t of the probability map, based on these observations. Since both the tissue types (we observe only the gray level) and their distribution are unknown, we need to use once more the EM-algorithm to estimate the proportions of the mixture. The EM algorithm maximizes the following quantity with respect to $P(Z_t = j)$:

$$\sum_{i=1}^{N} \ln \sum_{j=1}^{6} \frac{P(Z_t = j)}{\sqrt{2\pi\sigma^2_{i,j}}} \exp\left(-\frac{(x^{(i)}_{\phi_{\beta_i}(t)} - \mu_{i,j})^2}{2\sigma^2_{i,j}}\right). \tag{4}$$

We use the photometric parameters $(\mu_{i,j}, \sigma^2_{i,j})$ estimated previously on each one of the images of the training set. Figure 2 shows one sagittal slice of the tissue probability map. Notice that the anatomy is sharply described by the tissue probability map in the vicinity of the landmarks while it is more diffuse at longer distances.

4 Landmark Detection on a New Image

Consider a new image $(x_v)_{v \in V}$. Since the likelihood of the intensities is a function of β, the deformation parameters, we look for the deformation ϕ_β that maximizes the conditional likelihood $l(\beta)$. That is, $\hat{\beta} = \arg\max_\beta l(\beta)$ with

$$l(\beta) = \sum_{v \in V} \ln \sum_{j=1}^{6} g_j(x_v) P(Z_{\phi^{-1}(v)} = j). \tag{5}$$

The estimated landmark location is simply given by $\hat{\boldsymbol{L}} = \phi_{\hat{\beta}}(\boldsymbol{L}^*)$.

However, ϕ_β^{-1} cannot be expressed in a simple analytic form, so we make the change of variables $v = \phi_\beta(t)$. The likelihood becomes,

$$l(\beta) = \sum_{t \in \phi_\beta^{-1}(V)} \ln \sum_{j=1}^{6} g_j(x_{\phi_\beta(t)}) P(Z_t = j) |J_{\phi_\beta}(t)|, \tag{6}$$

with $|J_{\phi_\beta}(t)|$ the absolute value of the determinant of the Jacobian of the change of variable. We approximate it by computing the quantity on the tissue probability map support $\boldsymbol{T}$ instead of $\phi_\beta^{-1}(V)$. In addition, since the support of the deformation ϕ_β is

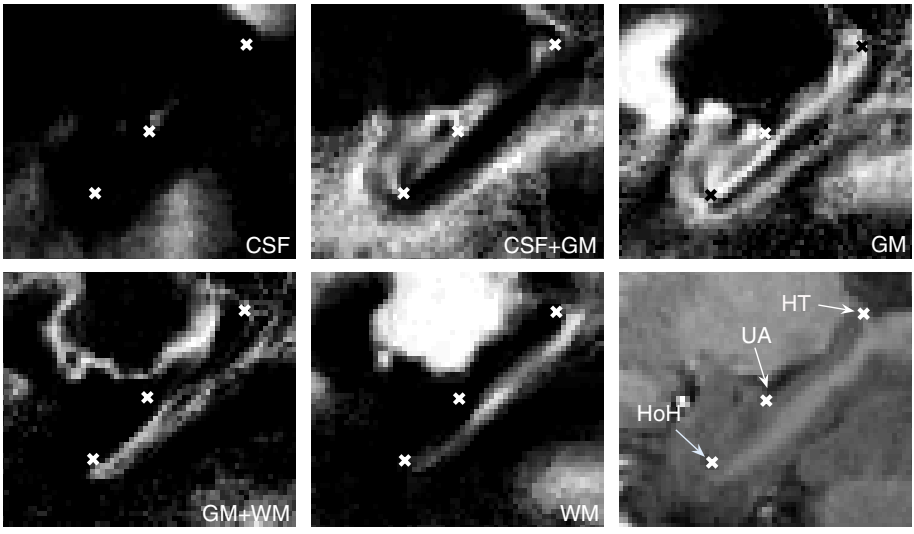


Fig. 2. Tissue probability map obtained when the deformation is estimated on three landmarks: HoH the head of the Hippocampus, HT the hippocampus tail and UA, the posterior apex of the hippocampal uncus. The slice corresponds to the sagittal section containing HoH, HT. UA is represented on the same slice although it lies in the next sagittal slice. We represent the 5 main tissues of the brain (the sixth one is not present in this region of the image). White regions have a high probability to belong to the corresponding tissue. The rightmost bottom image is an example of a registered image of the training set.

almost local, the computation can be reduced to the points around the landmarks, saving memory usage and time.

The derivative of the likelihood with respect to β can be computed exactly. The likelihood is maximized by simple gradient ascent in $\mathbb{R}^{3K}$. Conjugate gradient method did not improve the experimental results.

5 Experiments: Landmarking the Hippocampus in Brain MRI

The training set is composed of 38 T1-weighted MR brain images acquired on a Philips-Intera 3-Tesla scanner, with resolution 1mm^3. Brains were first manually transformed into standardized Talairach space using Analysis of Functional Neuroimages (AFNI) to provide a canonical orientation (anterior and posterior commissures (AC and PC) made co-linear) and approximate alignment. All the images have the same size after the transformation: $161 \times 191 \times 151$ voxels. An expert located the apex of the Head of the Hippocampus (HoH) and define on the same sagittal slice the Tail of the Hippocampus (HT) and he located the posterior apex of the hippocampal uncus (UA). (see Figure 2 bottom right)

We apply Algorithm 1 to predict simultaneously the location of HoH, HT and UA in the training set and in the testing set, composed of 9 images acquired in the same experimental setting as the training set and transformed into Talairach space. We exper-

imented with different values of σ but the results were comparable. We present in more details the case $\sigma_{HoH} = \sigma_{HT} = \sigma_{UA} = 5$.

After learning the model as described in section 2, we estimate the transformation parameters $\beta \in \mathbb{R}^9$ for each image independently. We assess the quality of the prediction by computing the Euclidean distance between the true landmarks $\boldsymbol{L}$ and the estimated position $\hat{\boldsymbol{L}}$. The hand-landmarking procedure defines HoH and HT so that they lie on the same sagittal slice. This is enforced in the optimization algorithm by expressing the constraints on β. For computational efficiency, we restrict the domain of computation to the set of voxels in the probability map within 3σ of each landmarks, since the gradient will be null at further distance of the landmarks. Table 1 presents the mean error over the images of the training and the testing set. For comparison,

Table 1. Mean prediction error in mm obtained by automatic landmarking for HoH, HT and UA on the training set (38 images) and on the testing set (9 images)

	mean error (mm) on the training set	mean error (mm) on the testing set
HoH	2.27 (1.58)	2.96 (1.17)
HT	2.49 (1.25)	2.57 (1.20)
UA	2.20 (1.30)	2.78 (1.59)

the specialist's intra-variability for HoH is 1.22mm ($\sigma = 0.92$), while the non-expert intra-variability on the same images for the same landmark is 3.58mm ($\sigma = 0.98$). The resulting inter-observer variability for HoH is 3.26mm ($\sigma = 0.98$). The algorithm reaches performance comparable to the inter-observers variability and offers for most of the images a reliable prediction of the landmark position.

Since locating landmarks in the image is aimed at reducing the local variability of the alignment, we compute the average gray-level across the images of the testing set and compare the initial alignment to the alignment after registration, using the automatic landmarks or the manual landmarks. For this experiment, we use the same registration technique as in section 3.2. Figure 3 presents the results. The improvement between (a) and (b) are concentrated around the landmarks. Around HT (top-right cross) the

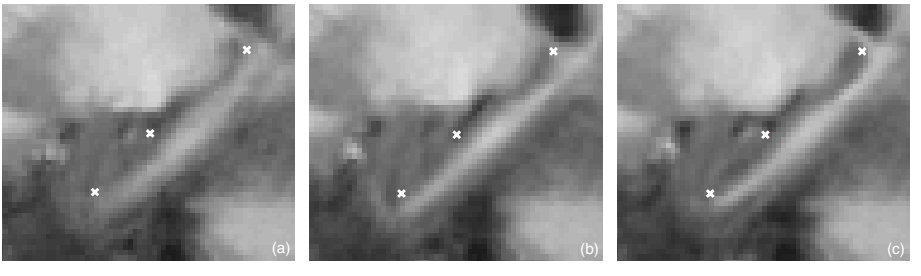


Fig. 3. Alignment experiment results on the testing set composed of 9 images initially aligned in the Talairach space. (a) represents the initial alignment without using landmarks, (b) represents the alignment obtained when the registration is made using the landmarks detected automatically, (c) alignment based on the expert's landmarks. The crosses represent HoH, HT and the projection of UA as it lies in the next sagittal slice.

average location of the tail is more consistent and accurate than in (a). The head of the hippocampus (bottom-left cross) is visible in (b) while it is not in (a). The overall alignment of the hippocampus is better in (b) than in (a).

6 Conclusion

We have proposed a simple and generic algorithm for automating the detection of landmarks on anatomical structures of the brain. The algorithm belongs to the class of generative models using a training set of manually located landmarks to specialize to the structure or region of interest. It adapts easily to various numbers and types of anatomical landmarks. Experimental results on brain MRI for locating three landmarks in the hippocampus are promising with an error comparable to the variability between landmarkers.

References

1. Dryden, I.L., Mardia, K.V.: Statistical Shape Analysis. Wiley (1998)
2. J.Talairach, Tournoux, P.: Co-planar stereotaxic Atlas of the Human Brain. Thieme Medical Publishers (1988)
3. Joshi, S., Miller, M.: Landmark matching via large deformation diffeomorphisms. In: IEEE Trans. in Image Processing. Volume 9. (2000) 1357–1370
4. Rohr, K., Stiehl, H., Sprengel, R., Buzug, T., Weese, J., Kuhn, M.: Landmark-based elastic registration using approximating thin-plate splines. IEEE Trans. Med. Img. **20** (2001) 526–534
5. Bookstein, F.: Morphometric Tools for Landmark Data: Geometry and Biology. Cambridge University Press (1992)
6. Thirion, J.P.: New feature points based on geometric invariants for 3d image registration. Int. J. of Computer Vision **18:2** (1996) 121–137
7. Wörz, S., Rohr, K.: Localization of anatomical point landmarks in 3d medical images by fitting 3d parametric intensity models. Medical Image Analysis **10** (2006) 41–58
8. Rohr, K.: Landmark-based Image Analysis using Geometric and Intensity Models. Kluwer Academic, Dordrecht (2001)
9. Ashburner, J., Friston, K.J.: Unified segmentation. NeuroImage (2005) 839–851
10. Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M.: Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron **33** (2002) 341–355
11. Wells, W., Kikinis, R., Grimson, W., Jolesz, F.: Adaptive segmentation of mri data. In: IEEE Trans. Med. Imag. Volume 15. (1996) 429–442
12. Dempster, A., Laird, N., Rubin, D.: Maximum likelihood from incomplete data via the em algorithm. J. R. Stat. Soc. **39** (1977) 1–38
13. 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** (1994) 192–205
14. Stark, C., Okado, Y.: Making memories without trying: Medial temporal lobe activity associated with incidental memory formation during recognition. J. of Neurosci. **23** (2003) 6748–6753
15. Camion, V., Younes, L.: Geodesic interpolating splines. In: Proceedings of Energy Minimisation in Computer Vision and Pattern Recognition (EMMCVPR). Volume 2134., Springer-Verlag (2001) 513–527

Affine and Deformable Registration Based on Polynomial Expansion

Gunnar Farnebäck and Carl-Fredrik Westin

Department of Radiology, Brigham and Womens' Hospital,
Harvard Medical School, Boston, MA 02115, USA

Abstract. This paper presents a registration framework based on the polynomial expansion transform. The idea of polynomial expansion is that the image is locally approximated by polynomials at each pixel. Starting with observations of how the coefficients of ideal linear and quadratic polynomials change under translation and affine transformation, algorithms are developed to estimate translation and compute affine and deformable registration between a fixed and a moving image, from the polynomial expansion coefficients. All algorithms can be used for signals of any dimensionality. The algorithms are evaluated on medical data.

1 Introduction

Medical images are brought into spatial correspondence, or aligned, by the use of registration algorithms. Nonrigid registration refers to the set of techniques that allow the alignment of datasets that are mismatched in a nonrigid, or nonuniform manner. Such misalignments can result from physical deformation processes, or can be a result of morphological variability. Nonrigid deformation is also characteristic of the organs and soft tissues of the abdomen and pelvis. In addition, nonrigid morphological differences can arise when comparisons are made among image datasets acquired from different individuals. These changes can be a result of normal anatomical variability or the product of pathological processes. Because the gross structure of the brain is essentially similar among humans (and even among related species), the factors described above tend to produce local nonrigid shape differences. Nonrigid brain registration techniques have numerous applications. They have been used to align scans of different brains, permitting the characterization of normal and pathological morphological variation (brain mapping). They have also been used to align anatomical templates with specific datasets, thus facilitating segmentation (i.e. segmentation by registration). More recently, these techniques have been used to capture changes which occur during neurosurgery. With the ongoing development of robust algorithms and advanced hardware platforms, further applications in surgical visualization and enhanced functional image analysis are inevitable.

Spatial normalization in clinical studies involves the registration of images and the generation of a stereotaxic atlas that represents the statistical distribution of the group at each voxel [1,2,3,4].

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 857–864, 2006.

2 Polynomial Expansion

The idea of polynomial expansion is to approximate the signal values in some neighborhood of each pixel with a polynomial. In this paper we are only interested in using linear or quadratic polynomials. In the quadratic case we have the local signal model, expressed in a local coordinate system,

$$f(\mathbf{x}) \sim \mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c, \tag{1}$$

where $\mathbf{A}$ is a symmetric matrix, $\mathbf{b}$ a vector and c a scalar. In the linear case we drop the first term and instead have the local signal model

$$f(\mathbf{x}) \sim \mathbf{b}^T\mathbf{x} + c. \tag{2}$$

The coefficients are determined from a weighted least squares fit to the signal values in the neighborhood. The weighting has two factors called certainty and applicability. These terms are the same as in normalized convolution [5]. The certainty is coupled to the signal values in the neighborhood. For example it is generally a good idea to set the certainty to zero outside the image. Then neighborhood points outside the image have no impact on the coefficient estimation.

The applicability determines the relative weight of points in the neighborhood based on their position in the neighborhood. Typically one wants to give most weight to the center point and let the weights decrease radially. In this paper we only use Gaussian applicability, both to simplify the presentation and because its Cartesian separability and isotropy are valuable features. This also allows us to describe the scale of the polynomial expansion with a single value, its standard deviation σ. This is directly related to the scale of the structures which will be captured by the expansion coefficients. Thus a small σ can be used to capture fine details and a large σ can be used to capture coarse details. More details about polynomial expansion can be found in [6].

3 Translation Estimation

For translation estimation we assume that the whole signal has been globally translated. As a preparation we start by analyzing what happens if a single quadratic polynomial is translated. First consider the quadratic polynomial $f_{\text{fixed}}(\mathbf{x})$ and construct a new signal f_{moving} by a global translation by $\mathbf{d}$,

$$f_{\text{fixed}}(\mathbf{x}) = \mathbf{x}^T\mathbf{A}_{\text{f}}\mathbf{x} + \mathbf{b}_{\text{f}}^T\mathbf{x} + c_{\text{f}}, \tag{3}$$

$$\begin{aligned} f_{\text{moving}}(\mathbf{x}) &= f_{\text{fixed}}(\mathbf{x}-\mathbf{d}) = (\mathbf{x}-\mathbf{d})^T\mathbf{A}_{\text{f}}(\mathbf{x}-\mathbf{d}) + \mathbf{b}_{\text{f}}^T(\mathbf{x}-\mathbf{d}) + c_{\text{f}} \\ &= \mathbf{x}^T\mathbf{A}_{\text{f}}\mathbf{x} + (\mathbf{b}_{\text{f}} - 2\mathbf{A}_{\text{f}}\mathbf{d})^T\mathbf{x} + \mathbf{d}^T\mathbf{A}_{\text{f}}\mathbf{d} - \mathbf{b}_{\text{f}}^T\mathbf{d} + c_{\text{f}} = \mathbf{x}^T\mathbf{A}_{\text{m}}\mathbf{x} + \mathbf{b}_{\text{m}}^T\mathbf{x} + c_{\text{m}}. \end{aligned} \tag{4}$$

Equating the quadratic and linear coefficients yields

$$\mathbf{A}_{\text{m}} = \mathbf{A}_{\text{f}}, \tag{5}$$

$$\mathbf{b}_{\text{m}} = \mathbf{b}_{\text{f}} - 2\mathbf{A}_{\text{f}}\mathbf{d}. \tag{6}$$

If $\mathbf{A}_\mathrm{f}$ is non-singular, (6) alone lets us solve for the translation $\mathbf{d}$, regardless of signal dimensionality,

$$\mathbf{d} = \frac{1}{2}\mathbf{A}_\mathrm{f}^{-1}(\mathbf{b}_\mathrm{f} - \mathbf{b}_\mathrm{m}). \tag{7}$$

In reality we do not have a single global polynomial but after polynomial expansion of the signal we get one polynomial approximation of the signal at each pixel. Let $\mathbf{A}_\mathrm{f}(\mathbf{x})$, $\mathbf{b}_\mathrm{f}(\mathbf{x})$, and $c_\mathrm{f}(\mathbf{x})$ be the quadratic polynomial expansion coefficients for the fixed image and $\mathbf{A}_\mathrm{m}(\mathbf{x})$, $\mathbf{b}_\mathrm{m}(\mathbf{x})$, and $c_\mathrm{m}(\mathbf{x})$ be the coefficients for the moving image. According to (5) we should have $\mathbf{A}_\mathrm{f}(\mathbf{x}) = \mathbf{A}_\mathrm{m}(\mathbf{x})$ but we cannot expect this to hold when the signal no longer is a single polynomial. Instead we use their average and introduce

$$\mathbf{A}(\mathbf{x}) = \frac{\mathbf{A}_\mathrm{f}(\mathbf{x}) + \mathbf{A}_\mathrm{m}(\mathbf{x})}{2}, \quad \Delta\mathbf{b}(\mathbf{x}) = \frac{1}{2}(\mathbf{b}_\mathrm{f}(\mathbf{x}) - \mathbf{b}_\mathrm{m}(\mathbf{x})) \tag{8}$$

to transform (6) into the primary constraint

$$\mathbf{A}(\mathbf{x})\mathbf{d} = \Delta\mathbf{b}(\mathbf{x}). \tag{9}$$

One possible approach now is to solve (9) pointwise and average the translation vectors over the image to estimate the global translation $\mathbf{d}$. This is not a good solution, however, since errors in $\mathbf{A}(\mathbf{x})$ and $\Delta\mathbf{b}(\mathbf{x})$ (caused by noise and model imperfections) become unnecessarily amplified in the pointwise computations. Instead we compute $\mathbf{d}$ by minimizing the squared error in the constraints over the whole image,

$$\epsilon^2 = \sum_{\mathbf{x}} \|\mathbf{A}(\mathbf{x})\mathbf{d} - \Delta\mathbf{b}(\mathbf{x})\|^2, \tag{10}$$

giving the least squares solution

$$\mathbf{d} = \Big(\sum \mathbf{A}(\mathbf{x})^T\mathbf{A}(\mathbf{x})\Big)^{-1} \sum \mathbf{A}(\mathbf{x})^T\Delta\mathbf{b}(\mathbf{x}). \tag{11}$$

3.1 Incorporating a Priori Knowledge

We started the derivation of (11) by considering a translation of a single polynomial. In that case (5) and (6) are valid regardless how large the translation is. For a real signal the differences in the polynomial coefficients at any given pixel between the fixed and the moving images are only partially explained by these equations. The second factor is that the polynomial approximations of the signal vary spatially as well. The larger the distance, the less similar the approximating polynomials can be expected to be. Thus the constraint (9) can be expected to be more uncertain the larger the translation.

Fortunately we are not restricted to comparing two polynomials at the same coordinate. If we happen to have some a priori knowledge about the translation we can instead compare the polynomial coefficients at $\mathbf{x}$ in the fixed image to the

coefficients at $\mathbf{x}+\tilde{\mathbf{d}}$ in the moving image, where $\tilde{\mathbf{d}}$ is the rounded a priori estimate. Then we effectively only need to estimate the relative translation between the true translation and $\tilde{\mathbf{d}}$, which hopefully is smaller. More precisely we can include this observation in the algorithm by replacing (8) by

$$\mathbf{A}(\mathbf{x}) = \frac{\mathbf{A}_\mathrm{f}(\mathbf{x}) + \mathbf{A}_\mathrm{m}(\tilde{\mathbf{x}})}{2}, \tag{12}$$

$$\boldsymbol{\Delta}\mathbf{b}(\mathbf{x}) = \frac{1}{2}(\mathbf{b}_\mathrm{f}(\mathbf{x}) - \mathbf{b}_\mathrm{m}(\tilde{\mathbf{x}})) + \mathbf{A}(\mathbf{x})\tilde{\mathbf{d}}(\mathbf{x}), \tag{13}$$

$$\tilde{\mathbf{x}} = \mathbf{x} + \tilde{\mathbf{d}}(\mathbf{x}). \tag{14}$$

We let $\tilde{\mathbf{d}}$ be space variant because we will need it later and also to allow us to avoid getting any $\tilde{\mathbf{x}}$ outside the border of the image.

3.2 Iterative and Multi-scale Translation Estimation

A consequence of the inclusion of an a priori translation estimate in the algorithm is that we can close the loop and iterate. A better a priori estimate means a smaller relative translation, which in turn improves the chances for a good new translation estimate. This can be done within one scale or between scales. In the first case we keep the polynomial expansion coefficients between the iterations and only recompute (14), (12), (13), and (11). In most cases this converges in only a few iterations, although not necessarily to a correct result if the translation is too large. Notice that since we are rounding the translation estimates at the start of each iteration we are very unlikely to get into any small scale oscillations. The initial translation estimate in the first step can be set to zero, unless actual knowledge about it is available.

3.3 Weighting the Constraints

As already has been discussed, the constraint (9) tends to become more uncertain the larger the translation is. One way to estimate this uncertainty is to see how much $\mathbf{A}_\mathrm{m}(\tilde{\mathbf{x}})$ differs from $\mathbf{A}_\mathrm{f}(\mathbf{x})$, since they according to (5) should ideally be equal. Thus we introduce

$$\gamma_2(\mathbf{x}) = e^{-\frac{\alpha\|\mathbf{A}_\mathrm{f}(\mathbf{x}) - \mathbf{A}_\mathrm{m}(\tilde{\mathbf{x}})\|_\mathrm{F}^2}{\|\mathbf{A}_\mathrm{f}(\mathbf{x})\|_\mathrm{F}^2 + \|\mathbf{A}_\mathrm{m}(\tilde{\mathbf{x}})\|_\mathrm{F}^2}}, \tag{15}$$

which will vary from 1 when $\mathbf{A}_\mathrm{m}(\tilde{\mathbf{x}})$ and $\mathbf{A}_\mathrm{f}(\mathbf{x})$ are equal down to $e^{-2\alpha}$ when they are opposite. The squared Frobenius norm $\|\cdot\|_\mathrm{F}^2$ is given by the sum of the squared matrix elements. Now we introduce $\gamma_2(\mathbf{x})$ as a weight in the squared error measure (10), by once more modifying the definition of $\mathbf{A}(\mathbf{x})$ and $\boldsymbol{\Delta}\mathbf{b}(\mathbf{x})$, replacing (12) and (13) by

$$\mathbf{A}(\mathbf{x}) = \gamma_2(\mathbf{x})\frac{\mathbf{A}_\mathrm{f}(\mathbf{x}) + \mathbf{A}_\mathrm{m}(\tilde{\mathbf{x}})}{2}, \tag{16}$$

$$\boldsymbol{\Delta}\mathbf{b}(\mathbf{x}) = \gamma_2(\mathbf{x})\left(\frac{1}{2}(\mathbf{b}_\mathrm{f}(\mathbf{x}) - \mathbf{b}_\mathrm{m}(\tilde{\mathbf{x}})) + \mathbf{A}(\mathbf{x})\tilde{\mathbf{d}}(\mathbf{x})\right). \tag{17}$$

3.4 Using Linear Polynomial Expansion

We can repeat these derivations using the linear polynomial model (2), instead of the quadratic model (1). This changes (3) and (4) into

$$f_{\text{fixed}}(\mathbf{x}) = \mathbf{b}_{\text{f}}^T\mathbf{x} + c_{\text{f}} \tag{18}$$

$$f_{\text{moving}}(\mathbf{x}) = f_{\text{fixed}}(\mathbf{x} - \mathbf{d}) = \mathbf{b}_{\text{f}}^T\mathbf{x} + c_{\text{f}} - \mathbf{b}_{\text{f}}^T\mathbf{d} = \mathbf{b}_{\text{m}}^T\mathbf{x} + c_{\text{m}}. \tag{19}$$

Equating the coefficients now gives

$$\mathbf{b}_{\text{m}} = \mathbf{b}_{\text{f}}, \quad c_{\text{m}} = c_{\text{f}} - \mathbf{b}_{\text{f}}^T\mathbf{d}, \tag{20}$$

and the correspondences to (15)–(17) become

$$\gamma_1(\mathbf{x}) = e^{-\frac{\alpha\|\mathbf{b}_{\text{f}}(\mathbf{x}) - \mathbf{b}_{\text{m}}(\tilde{\mathbf{x}})\|_{\text{F}}^2}{\|\mathbf{b}_{\text{f}}(\mathbf{x})\|_{\text{F}}^2 + \|\mathbf{b}_{\text{m}}(\tilde{\mathbf{x}})\|_{\text{F}}^2}}, \tag{21}$$

$$\mathbf{b}(\mathbf{x}) = \gamma_1(\mathbf{x})\frac{\mathbf{b}_{\text{f}}(\mathbf{x}) + \mathbf{b}_{\text{m}}(\tilde{\mathbf{x}})}{2}, \tag{22}$$

$$\Delta c(\mathbf{x}) = \gamma_1(\mathbf{x})(\mathbf{c}_{\text{f}}(\mathbf{x}) - \mathbf{c}_{\text{m}}(\tilde{\mathbf{x}}) + \mathbf{b}(\mathbf{x})^T\tilde{\mathbf{d}}(\mathbf{x})). \tag{23}$$

Now we add the linear constraint into (10), using β_1 and β_2 as weights for the linear and quadratic parts,

$$\epsilon^2 = \sum_{\mathbf{x}} \beta_1(\mathbf{b}(\mathbf{x})^T\mathbf{d} - \Delta c(\mathbf{x}))^2 + \beta_2\|\mathbf{A}(\mathbf{x})\mathbf{d} - \boldsymbol{\Delta}\mathbf{b}(\mathbf{x})\|^2, \tag{24}$$

with the least squares solution

$$\mathbf{G} = \sum \beta_1\mathbf{b}(\mathbf{x})\mathbf{b}(\mathbf{x})^T + \beta_2\mathbf{A}(\mathbf{x})^T\mathbf{A}(\mathbf{x}), \tag{25}$$

$$\mathbf{h} = \sum \beta_1\mathbf{b}(\mathbf{x})\Delta c(\mathbf{x}) + \beta_2\mathbf{A}(\mathbf{x})^T\boldsymbol{\Delta}\mathbf{b}(\mathbf{x}), \tag{26}$$

$$\mathbf{d} = \mathbf{G}^{-1}\mathbf{h}. \tag{27}$$

4 Affine Registration

Now we turn to the more interesting case of affine registration and to simplify the presentation we do this in 2D. Generalizations to higher dimensionalities are straightforward. The deformation field is now space variant

$$\begin{aligned} d_x(x, y) &= a_1x + a_2y + a_5, \\ d_y(x, y) &= a_3x + a_4y + a_6, \end{aligned} \tag{28}$$

and we can rewrite this in matrix form as

$$\mathbf{d}(\mathbf{x}) = \mathbf{S}(\mathbf{x})\mathbf{p}, \tag{29}$$

$$\mathbf{S}(\mathbf{x}) = \begin{pmatrix} x & y & 0 & 0 & 1 & 0 \\ 0 & 0 & x & y & 0 & 1 \end{pmatrix}, \tag{30}$$

$$\mathbf{p} = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \end{pmatrix}^T. \tag{31}$$

To estimate the parameters of the affine deformation field is as simple as combining (24) and (29),

$$\epsilon^2 = \sum_{\mathbf{x}} \beta_1(\mathbf{b}(\mathbf{x})^T\mathbf{S}(\mathbf{x})\mathbf{p} - \Delta c(\mathbf{x}))^2 + \beta_2\|\mathbf{A}(\mathbf{x})\mathbf{S}(\mathbf{x})\mathbf{p} - \boldsymbol{\Delta}\mathbf{b}(\mathbf{x})\|^2 \tag{32}$$

and we get the parameters $\mathbf{p}$ from the least squares solution

$$\mathbf{G} = \sum \beta_1\mathbf{S}_i^T\mathbf{b}_i\mathbf{b}_i^T\mathbf{S}_i + \beta_2\mathbf{S}_i^T\mathbf{A}_i^T\mathbf{A}_i\mathbf{S}_i, \tag{33}$$

$$\mathbf{h} = \sum \beta_1\mathbf{S}_i^T\mathbf{b}_i\Delta c_i + \beta_2\mathbf{S}_i\mathbf{A}_i^T\boldsymbol{\Delta}\mathbf{b}_i, \tag{34}$$

$$\mathbf{p} = \mathbf{G}^{-1}\mathbf{h}, \tag{35}$$

where we have replaced the spatial coordinate by a subscript to shorten the notation. We can notice that this approach not only works for affine deformation but for any parametric deformation field which is linear in its parameters.

5 Deformable Registration

We can derive a deformable registration algorithm if we relax the assumption about a global translation or a globally affine deformation, and instead do the summation in either of (24) or (32) over a neighborhood around each pixel, thereby getting an estimate for each pixel. For computational reasons we will only work with local translations. This does not impose a strong structure on the deformation field, but due to lowpass filtering in the algorithm, the deformation field automatically becomes smooth. More exactly we change (24) to

$$\epsilon^2(\mathbf{x}) = \sum_{\mathbf{y}} w(\mathbf{y})(\beta_1(\mathbf{b}(\mathbf{x}-\mathbf{y})^T\mathbf{d}(\mathbf{x}) - \Delta c(\mathbf{x}-\mathbf{y}))^2 + \beta_2\|\mathbf{A}(\mathbf{x}-\mathbf{y})\mathbf{d}(\mathbf{x}) - \boldsymbol{\Delta}\mathbf{b}(\mathbf{x}-\mathbf{y})\|^2), \tag{36}$$

where w weights the points in the neighborhood around each pixel. This weight can be any lowpass function but here we assume it to be Gaussian. Clearly this equation can be interpreted as a convolution of the pointwise contributions to the squared error (24) with the lowpass filter w. The solution is

$$\mathbf{G}(\mathbf{x}) = \beta_1\mathbf{b}(\mathbf{x})\mathbf{b}(\mathbf{x})^T + \beta_2\mathbf{A}(\mathbf{x})^T\mathbf{A}(\mathbf{x}), \tag{37}$$

$$\mathbf{h}(\mathbf{x}) = \beta_1\mathbf{b}(\mathbf{x})\Delta c(\mathbf{x}) + \beta_2\mathbf{A}(\mathbf{x})^T\boldsymbol{\Delta}\mathbf{b}(\mathbf{x}), \tag{38}$$

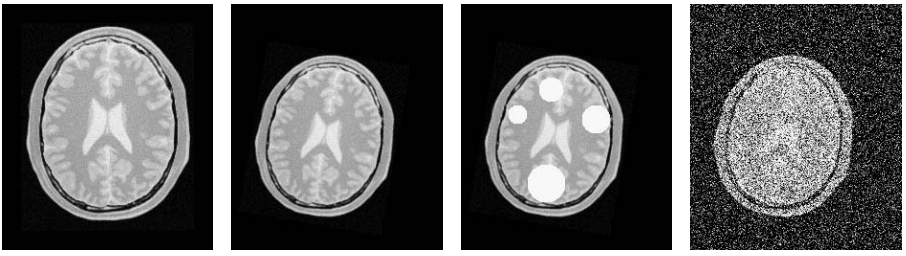
$$\mathbf{G}_{\text{avg}}(\mathbf{x}) = (\mathbf{G} * w)(\mathbf{x}), \tag{39}$$

$$\mathbf{h}_{\text{avg}}(\mathbf{x}) = (\mathbf{h} * w)(\mathbf{x}), \tag{40}$$

$$\mathbf{d}(\mathbf{x}) = \mathbf{G}_{\text{avg}}(\mathbf{x})^{-1}\mathbf{h}_{\text{avg}}(\mathbf{x}). \tag{41}$$

6 Experiments

The affine and deformable registration algorithms have been implemented in Matlab and in C++ using the Insight Toolkit (ITK) [7]. Both implementations have similar performance and have been tested on a Pentium IV at 2.8 GHz. The time measurements in this section are for the C++ implementation.



(a) original (b) transformed (c) blobs inserted (d) noise added

Fig. 1. Brain images from ITK (a, b) and corrupted versions (c, d)

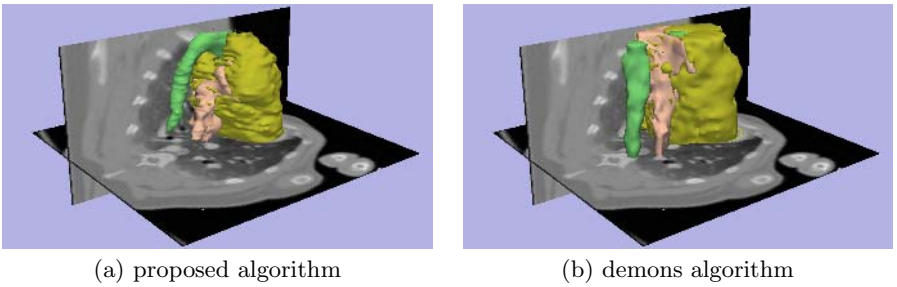
6.1 Affine Registration

The affine registration has been tested in 2D on example data from ITK. Figure 1(a) shows a brain image. This has been translated, rotated by 10° and scaled down by a factor 1.2 in Fig. 1(b). The affine registration algorithm takes 3.0 seconds and gives affine parameters corresponding to a rotation of 9.998° and a minimum and maximum scaling of 1.1991 and 1.1999 respectively. As a comparison the ITK example code for affine registration, `ImageRegistration9`, takes 28 seconds and gives a rotation estimate of 9.68° and scalings 1.1994 and 1.2023.

To test the sensitivity to disturbances, we have inserted a number of blobs in the second image, giving the image shown in Fig. 1(c). In Fig. 1(d) we have the result of adding a substantial amount of (truncated) Gaussian noise. Our algorithm for (c) takes 3.0 seconds and gives a rotation angle of 9.989° and for (d) it takes 4.5 seconds and gives an angle of 10.24°. The ITK example code takes 35 seconds and gives an angle of 9.22° for (c) and 53 seconds and 9.08° for (d). The scaling parameters are at most off by 0.5% in all cases. The differences in time is a question of a varying number of iterations for both methods.

6.2 Deformable Registration

The deformable registration algorithm has been tested in 3D on pig CT data (spatial resolution 1x1x3 mm) in a segmentation by registration experiment. This



(a) proposed algorithm (b) demons algorithm

Fig. 2. Warped pig organs

involves one data set where a few organs have been hand-segmented and a second data set where no segmentation is known. The CT volumes were registered using the deformable registration algorithm to obtain a deformation field, which was then used to warp the segmentation over from the first data set to the second. Figure 2(a) shows the warped segmentations superimposed on the second volume. For comparison Fig. 2(b) shows the corresponding result when the ITK implementation of demons algorithm (example code `DeformableRegistration2` converted to 3D) was used. The anatomy is poorly aligned in the latter case and it is clear that our algorithm performed more robustly than the demons algorithm in this example. The deformable registrations were preceded by affine registration with corresponding parameters.

7 Conclusions

We have presented a new registration framework and derived algorithms for translation estimation, affine registration, and deformable registration. All algorithms are valid for any dimensionality. Experiments on medical data indicate that the affine registration is both fast and accurate, and that the deformable registration performs robustly.

Acknowledgments

This work was supported by grants from the DARPA, executed by the U.S. Army Medical Research and Materiel Command/TATRC Cooperative Agreement, Contract # W81XWH-04-2-0012, and NIH P41-RR13218.

References

1. Mazziotta, J.C., Toga, A.W., Evans, A.C., Fox, P., Lancaster, J.L.: A probabilistic atlas of the human brain: Theory and rationale for its development. Neuroimage **2**(2) (1995) 89–101
2. Thompson, P.M., Toga, A.W.: Detection, visualization and animation of abnormal anatomic structure with a deformable probabilistic atlas based on random vector field transformations. Medical Image Analysis **1**(4) (1997) 271–294
3. Grenander, U., Miller, M.: Computational anatomy: an emerging discipline. Q. Appl. Math. **56**(4) (1998) 617–694
4. Guimond, A., Meunier, J., Thirion, J.P.: Average brain models: a convergence study. Comput. Vision Imag. Understand. **77**(2) (2000) 192–210
5. Knutsson, H., Westin, C.F.: Normalized and differential convolution: Methods for interpolation and filtering of incomplete and uncertain data. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York City, USA, IEEE (1993) 515–523
6. Farnebäck, G.: Polynomial Expansion for Orientation and Motion Estimation. PhD thesis, Linköping University, SE-581 83 Linköping, Sweden (2002) Dissertation No 790, ISBN 91-7373-475-6.
7. ITK: The Insight Segmentation and Registration Toolkit, `http://www.itk.org`

Simultaneous Multiple Image Registration Method for T_1 Estimation in Breast MRI Images

Jonathan Lok-Chuen Lo[1], Michael Brady[1], and Niall Moore[2]

[1] Wolfson Medical Vision Laboratory, University of Oxford, Oxford, UK
{jlo, jmb}@robots.ox.ac.uk
[2] MRI unit, John Radcliffe Hospital, Oxford, UK

Abstract. The estimation and subsequent use of tissue $T_1(\mathbf{x})$ parameters at each image location $\mathbf{x}$ can potentially lead to a more reliable classification of breast tissues. T_1 values can be estimated using multiple (typically 3) MRI images of different flip angles. However, breathing and other slight movements can render the highly non-linear estimation procedure error-prone. In this paper, a simultaneous multiple image registration method is proposed to solve this problem. The registration method is built upon the idea of conserving inverse consistency and transitivity among the multiple image transformations. The algorithm is applied to both simulated data and real breast MRI images. The performance is compared with existing pairwise image registration method. The results clearly indicate that the simultaneous multiple image registration algorithm leads to much more accurate T_1 estimation.

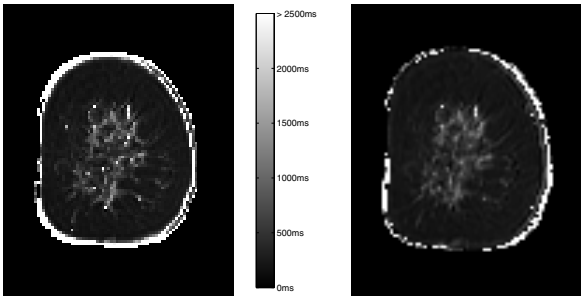
1 Introduction

Contrast-enhanced magnetic resonance imaging (CE-MRI) has gained considerable attention over recent years for early detection of breast cancers, and is increasingly used for cancer staging and treatment planning. This imaging modality is particularly useful for younger women, for whom the dense radiographic density of glandular tissues often render X-ray mammography ineffective.

Due to the frequent lack of contrast difference between abnormalities and normal tissues in conventional breast MRI, contrast agents such as gadolinium diethylene triamine pentaacetic acid (Gd-DTPA) are widely used to enable the detection and characterization of tumours, because they highlight areas with increased vascular density, which correlate with the presence of tumours. Temporal scans of the breast (approximately every one minute) are made before, during and after the administration of the contrast agent. The dynamic MRI image sequences are often fitted using a pharmacokinetic model to provide a physiologically relevant characterization of the underlying tissues [1]. Such curves/parameters have been applied clinically to identify and characterize tumours into malignant or benign classes, although the success has been variable with generally very good sensitivity ($> 95\%$) but often poor specificity [2].

A major reason for the poor specificity stems from the erroneous assumption that intensity enhancement is linearly proportional to contrast agent concentration. In [3], the authors suggested that the relationship is both non-linear and

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 865–872, 2006.



(a) (b)

Fig. 1. Estimated T_1 mapping for breast MRI images, with (a) no registration and (b) pairwise nonrigid registration. The near white pixel indicates $T_1 > 2500$ ms, which is unrealistically large for breast tissues.

highly dependent on the intrinsic tissue T_1 values, and that contrast agent concentration is to a good approximation linear proportional to the change in T_1. It follows that accurate estimation of T_1 at each voxel is a necessary prerequisite for more reliable tissue classification using the uptake curves of the pharmacokinetic model. T_1 values may, for example, be estimated from images resulting from using the fast spoiled gradient echo (FSPGR) pulse sequence. The MRI signal then takes the form [4]:

$$S(t) = g\rho \exp\left(-T_E/T_2(t)\right) \sin\alpha \frac{1 - \exp\left(-T_R/T_1(t)\right)}{1 - \cos\alpha \exp\left(-T_R/T_1(t)\right)}, \tag{1}$$

where S is the measured MR signal, g is the scanner system gain, ρ is the proton density, T_E is the echo time and T_R is the repetition time and α is the flip angle. A series of n volumes, each with a different flip angle α_n, are acquired. The model in (1) is fitted to $S_n(0)$ vs α_n data at each voxel to obtain the unknown parameters $T_{10} = T_1(0)$ and k, where $k = g\rho \exp\left(-T_E/T_2(t)\right)$.

Armitage et al. [3] used Monte Carlo simulation to calculate the optimal number of flip angles required for measuring the ranges of T_1 found in the breast. They concluded that the use of three flip angles ($\{3°, 10°, 17°\}$ were suggested) provides the best compromise between acquisition time and the accuracy of the resulting T_1 estimation.

In practice, however, breathing and other slight positional movements result in large errors in the T_1 estimation. For example, in Figure 1(a), the effect of this misalignment is shown. It can be seen that unrealistically large T_1 values (> 2500 ms) are estimated around the edge of the breast, shown as white pixels in the image. To address this problem, we first used the non-rigid registration algorithm, on a pairwise image basis, described in [5], i.e. two separate registrations are performed on two pairs of images. However, we found that the results are unpromising (see Figure 1(b)).

In this paper, we propose a new method to simultaneously register three MRI images of different flip angles, based on the idea of constraining the inverse consistency and transitivity among the transformations, as proposed in [6]. Following

Rueckert, B-splines are used as the underlying non-rigid transformations, with mutual information as the similarity measure. Though alternative non-rigid registration algorithms could have been used, this provided a more direct test of the added value of the consistency and transitivity constraint. Simulated data, as well as real MRI breast images, are tested using our proposed technique. The results show that the method significantly improves the accuracy of the estimated T_1 values.

2 Method

Christensen and Johnson suggested the idea of constraining inverse consistency and transitivity for multiple images registration [6]. A pair of transformations from image I_1 to I_2 and I_2 to I_1 are said to be inverse consistent if the composition of the two transformations is the identity mapping. A set of transformations is said to be transitive if the composition of transformations from I_1 to I_2 and I_2 to I_3 equals to the transformation from I_1 to I_3. In [6], the registration is based on the idea of fluid flow, where the smoothness of the transformation is guaranteed by using a linear-elastic regularization model. However, the complexity of the transformation requires large magnitude nonlinear kinematics, and significant computation time is needed. Furthermore, gray levels are used to drive the similarity measure, which is not suitable in our application where intensities are vastly different among images with different flip angles.

Based on the above ideas, Geng et al. [7] had proposed a contour/surface-based method, where three manifolds were registered at the same time, and the results seem encouraging. Closest point distance was used as the cost function for minimization. However, in our application, the tissue boundaries are not always clearly defined, especially for images with low flip angle where signal contrast between tissue types tends to be fairly low, and therefore contours cannot be found accurately. Our new method is proposed in the following sections.

2.1 Multiple Image Registration

Assume that all images are of size $X \times Y = N$, and let us denote the discrete position of the pixels as $\mathbf{r}(i) \in \Omega_d$, where $\Omega_d = \{(x, y) : x = 1, 2, \ldots, X;\ y = 1, 2, \ldots, Y\}$ and $i \in \{1, 2, \ldots, N\}$. A set of transformations is defined as $\mathcal{T} = \{\mathbf{T}_{ab} : 1 \leq a,\ b \leq 3,\ a \neq b\}$, where $\mathbf{T}_{ab}$ indicates the transformation mapping any point in image I_a to the corresponding point in image I_b, i.e. $\mathbf{T}_{ab} : (x_a, y_a) \mapsto (x_b, y_b)$. The goal of registration is to find the transformation set $\mathcal{T}$ such that the following cost function is minimized:

$$C(\mathcal{T}) = \alpha\, C_{\text{similarity}} + \beta\, C_{\text{smooth}} + \gamma\, C_{\text{invertibility}} + \nu\, C_{\text{transitivity}}, \qquad (2)$$

where $C_{\text{similarity}}$ is the image similarity cost function, C_{smooth} defines the constraint of the transformation smoothness, $C_{\text{invertibility}}$ and $C_{\text{transitivity}}$ are the penalty terms for the inverse consistency and transitivity respectively.

Tensor products of B-splines [5] are used to represent the transformation, with $\Phi = \{(\mu, \nu) : \mu = 1, 2, \ldots, n_x;\ \nu = 1, 2, \ldots, n_y\}$ denoting the underlying $n_x \times n_y$

mesh of control points. The 2-D transformation can be written as the tensor product of cubic splines:

$$\mathbf{T}(x,y) = \sum_{m=0}^{3}\sum_{n=0}^{3} B_m(u)B_n(v)\phi_{i+m,j+n}, \tag{3}$$

where $i = \lfloor x/n_x \rfloor - 1, j = \lfloor y/n_y \rfloor - 1, u = x/n_x - i - 1, v = y/n_y - j - 1$ and B_l represents the lth basis function of the B-spline. Registration becomes the problem of finding the set of control points Φ which minimizes the cost function in (2).

Similarity Cost Function. As the difference in contrast between the three MRI images with different flip angles is quite pronounced, direct comparison of image intensities is not appropriate. Mutual information is chosen to be the similarity measure. The cost function is defined as:

$$C_{\text{similarity}} = \sum_{a=1}^{3}\sum_{\substack{b=1\\b\neq a}}^{3} \Big[H(I_a) + H(\mathbf{T}_{ba}(I_b)) - H(I_a, \mathbf{T}_{ba}(I_b))\Big], \tag{4}$$

where $H(I_i)$ denotes the marginal entropy of image i and $H(I_i, I_j)$ is the joint entropy of image i and j.

Smoothness Constraint. The deformation of the breast tissues is assumed to be smooth, and this can be enforced on the B-spline transformation by including a penalty term which associates the curvature of the splines:

$$C_{\text{smooth}} = \sum_{a=1}^{3}\sum_{\substack{b=1\\b\neq a}}^{3}\sum_{i=1}^{N} \left[\left(\frac{\partial^2 \mathbf{T}_{ab}(\mathbf{r}(i))}{\partial x^2}\right)^2 + \left(\frac{\partial^2 \mathbf{T}_{ab}(\mathbf{r}(i))}{\partial y^2}\right)^2 + 2\left(\frac{\partial^2 \mathbf{T}_{ab}(\mathbf{r}(i))}{\partial x \partial y}\right)^2\right]. \tag{5}$$

This term corresponds to the bending energy of a thin-plate, which acts as a constraint on the smoothness of the transformation.

Inverse Consistency Constraint. To ensure that two transformations are inverse consistent, we want the combined transformation, $\mathbf{T}_{ab} \circ \mathbf{T}_{ba}$, as close to identity transformation as possible. Therefore, the inverse consistency constraint is defined as the following:

$$C_{\text{invertibility}} = \sum_{a=1}^{3}\sum_{\substack{b=1\\b\neq a}}^{3}\sum_{i=1}^{N} \left\|\mathbf{r}(i) - \mathbf{T}_{ab} \circ \mathbf{T}_{ba}(\mathbf{r}(i))\right\|^2. \tag{6}$$

Transitivity Constraint. Similarly, transitivity describes how close the combined transformation, $\mathbf{T}_{ab} \circ \mathbf{T}_{bc}$, is compared to the direct transformation $\mathbf{T}_{ac}$. We define the transitivity constraint to be:

$$C_{\text{transitivity}} = \sum_{a=1}^{3} \sum_{\substack{b=1 \\ b\neq a}}^{3} \sum_{\substack{c=1 \\ c\neq a \\ c\neq b}}^{3} \sum_{i=1}^{N} \left\| \mathbf{T}_{ac}(\mathbf{r}(i)) - \mathbf{T}_{ab} \circ \mathbf{T}_{bc}(\mathbf{r}(i)) \right\|^2 . \tag{7}$$

2.2 Implementation

The registration method is implemented using ITK[8]. In order to provide an efficient registration method, a gradient-based optimization technique is used, and hence the gradient vector of the cost function, i.e. $\nabla C = \partial C(\Phi)/\partial \Phi$, is required.

The gradient of the inverse consistency constraints with respect to each transformation parameters is:

$$\frac{\partial C_{\text{invertibility}}}{\partial \phi_{\mu,\nu}} = -2 \sum_{a=1}^{3} \sum_{\substack{b=1 \\ b\neq a}}^{3} \sum_{i=1}^{N} \left[\left\| \mathbf{r}(i) - \mathbf{T}_{ab} \circ \mathbf{T}_{ba}(\mathbf{r}(i)) \right\| \cdot \left\| \frac{\partial \mathbf{r}'(i)}{\partial \phi_{\mu,\nu}} \right\| \right], \tag{8}$$

where

$$\mathbf{r}'(i) = \begin{cases} \mathbf{T}_{ba}(\mathbf{r}(i)) & \text{if } \phi_{\mu,\nu} \text{ in } \mathbf{T}_{ba} \\ \mathbf{T}_{ab} \circ \mathbf{T}_{ba}(\mathbf{r}(i)) & \text{if } \phi_{\mu,\nu} \text{ in } \mathbf{T}_{ab}. \end{cases} \tag{9}$$

Similarly, the gradient of the transitivity constraint is:

$$\frac{\partial C_{\text{transitivity}}}{\partial \phi_{\mu,\nu}} = -2k \sum_{a=1}^{3} \sum_{\substack{b=1 \\ b\neq a}}^{3} \sum_{\substack{c=1 \\ c\neq a \\ c\neq b}}^{3} \sum_{i=1}^{N} \left[\left\| \mathbf{T}_{ac}(\mathbf{r}(i)) - \mathbf{T}_{ab} \circ \mathbf{T}_{bc}(\mathbf{r}(i)) \right\| \cdot \left\| \frac{\partial \mathbf{r}'(i)}{\partial \phi_{\mu,\nu}} \right\| \right], \tag{10}$$

where

$$\mathbf{r}'(i) = \begin{cases} \mathbf{T}_{bc}(\mathbf{r}(i)) & \text{if } \phi_{\mu,\nu} \text{ in } \mathbf{T}_{bc} \\ \mathbf{T}_{ab} \circ \mathbf{T}_{bc}(\mathbf{r}(i)) & \text{if } \phi_{\mu,\nu} \text{ in } \mathbf{T}_{ab} \\ \mathbf{T}_{ac}(\mathbf{r}(i)) & \text{if } \phi_{\mu,\nu} \text{ in } \mathbf{T}_{ac} \end{cases} \tag{11}$$

$$k = \begin{cases} 1 & \text{if } \phi_{\mu,\nu} \text{ in } \mathbf{T}_{bc} \text{ or } \mathbf{T}_{ab} \\ -1 & \text{if } \phi_{\mu,\nu} \text{ in } \mathbf{T}_{ac}. \end{cases} \tag{12}$$

3 Results

Our proposed registration method was tested by using both simulated data and real clinical MRI images. Simulated MRI images of flip angles $\{3°, 10°, 17°\}$ were generated from a phantom T_1 image. The simulated images were then distorted by using randomly generated B-spline deformation field places upon the images,

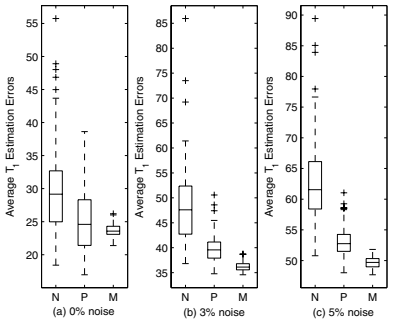


Fig. 2. Box plots for the 100 simulated images for (a) 0%, (b) 3% and (c) 5% noise are shown. The symbol N, P and M represent the results for: No registration, Pairwise registration and Multiple image registration. The boxes shown stretch from the 25th percentile (bottom) to the 25th percentile (top), and the median is shown as the line across the box.

with Gaussian noise being added to the final images. A hundred sets of such images were generated with 0%, 3% and 5% noise level. The images were then registered using pairwise non-rigid registration and our proposed simultaneous multiple image registration method, with control points spacing $\delta = 10$ pixels. The pairwise registration uses the same cost function in (2), but with both γ and ν set to zero. Due to the intrinsic smoothness properties of B-splines, the choice of λ is not critical for our application, and $\lambda = 0.01$ is chosen.

T_1 values were then estimated and compared to the known values of the phantom image. The average errors in the estimated T_1 per pixel are plotted in Figure 2. Our multiple image registration has definitely improved the T_1 estimation from images without any registration. Our method also outperforms the normal pairwise registration, and more importantly, the error ranges are greatly reduced.

Tests have also been carried out on clinical breast MRI images. Data from eight patients were used. All volumes were acquired on General Electric Medical Systems (Genesis Signa) 1.5 T clinical scanner using a T1-weighted 3D fast spoiled gradient echo (FSPGR) pulse sequence ($TE = 4.2\,\text{ms}, TR = 8.9\,\text{ms}$). Three image volumes were acquired with flip angles set to $\{3°, 10°, 17°\}$. Each image volume contains between 60-80 images. The slice is coronal orientation, with 3 mm spacing between slices.

Table 1. The average residual errors in the T_1 estimation for the eight sets of breast MRI images

Example number	No registration	Pairwise registration	Multiple image registration
1	12.3132	8.0006	5.5265
2	9.7924	5.0489	3.9482
3	7.5746	3.8106	2.0601
4	18.3493	7.8637	6.0015
5	15.9211	9.3463	7.1164

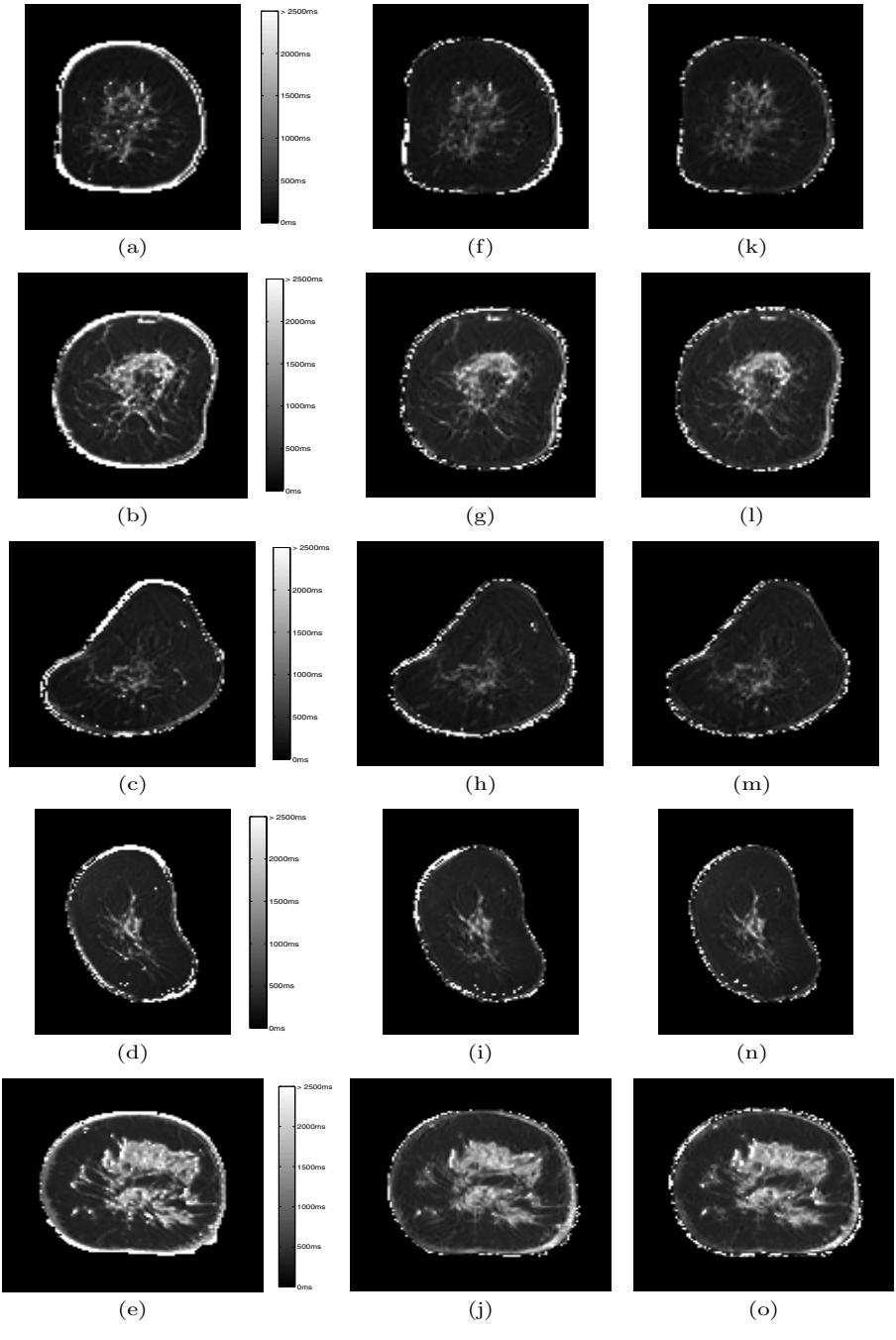


Fig. 3. Some examples of estimated T_1 using (a)-(e) no registration, (f)-(j) pairwise registration and (k)-(o) multiple image registration. The color scheme is similar to that in Figure 1, with near white pixel representing high T_1 values, i.e. > 2500 ms.

Similarly, these image sets were registered using both the pairwise registration and our multiple image registration. The registered images were used to estimate the T_1 values. Because, in this case we do not have the ground truth T_1 values, the residual error in our T_1 values estimation were recorded. Some examples of the estimated T_1 mapping are shown in Figure 3, and their average residual errors are shown in Table 1.

As can be seen from the estimated T_1 mapping for images without registration, regions of high estimated T_1 values (> 2500 ms, as indicated by near white pixels in the images) are mainly concentrated in the breast boundaries, as expected. Both registration techniques improve the results, but it clearly shows that our multiple image registration method performs better. The unrealistic high T_1 values are almost completely removed. With the simultaneous multiple image registration, we have shown that better estimation of T_1 values is achieved.

4 Discussion

In this paper, we have proposed a simultaneous multiple image registration technique for aligning multiple MRI images taken at different flip angles. The aim of the registration is to enable a better estimation of the pre-injection T_1. The initial testing on clinical data shows that the registration has indeed improved the T_1 estimation. Furthermore, our proposed multiple image registration has also outperformed the normal pairwise registration method.

The ultimate goal of our research, however, is to show that with improved T_1 estimation, we can achieve a more reliable classification of breast tissues from CE-MRI images. Further work is currently underway for such study.

References

1. Tofts, P.S.: Modeling tracer kinetics in dynamic Gd-DTPA MR imaging. Journal of Magnetic Resonance Imaging **7** (1997) 91–101
2. Rankin, S.: Mri of the breast. British Journal of Radiology **73** (2000) 806818
3. Armitage, P., Behrenbruch, C., Brady, M., Moore, N.: Extracting and visualizing physiological parameters using dynamic contrast-enhanced magnetic resonance imaging of the breast. Medical Image Analysis **9** (2005) 315–329
4. Christen, K.A., Grant, D.M., Schulman, E.M., Walling, C.: Optimal determination of relaxation-times of Fourier-transform nuclear magnetic-resonance. Journal of Physical Chemistry **78** (1974) 1971–1977
5. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J.: Non-rigid registration using free-form deformations: Application to breast MR images. IEEE Transactions on Medical Imaging **18** (1999) 712–721
6. Christensen, G.E., Johnson, H.J.: Consistent image registration. IEEE Transactions on Medical Imaging **20** (2001) 568–582
7. Geng, X.J., Kumar, D., Christensen, G.E.: Transitive inverse-consistent manifold registration. In: Information Processing in Medical Imaging, Proceedings. Volume 3565 of Lecture Notes in Computer Science. (2005) 468–479
8. Insight Toolkit, ITK: http://www.itk.org. (2005)

New CTA Protocol and 2D-3D Registration Method for Liver Catheterization

Martin Groher[1], Nicolas Padoy[1], Tobias F. Jakobs[2], and Nassir Navab[1]

[1] Chair for Computer Aided Medical Procedures (CAMP), Technische Universität München, Germany
{groher, padoy, navab}@cs.tum.edu
[2] Institute for Clinical Radiology, University of Munich, Grosshadern Hospital
tobias.jakobs@med.uni-muenchen.de

Abstract. 2D-3D registration for angiographic liver interventions is an unsolved problem mainly because of two reasons. First, a suitable protocol for Computed Tomography Angiography (CTA) to contrast liver arteries is not used in clinical practice. Second, an adequate registration algorithm which addresses the difficult task of aligning deformed vessel structures has not been developed yet. We address the first issue by introducing an *angiographic* CT scanning phase and thus create a strong link between radiologists and interventionalists. The scan visualizes arteries similar to the vasculature captured with an intraoperative C-arm acquiring Digitally Subtracted Angiograms (DSAs). Furthermore, we propose a registration algorithm using the new CT phase that aligns arterial structures in two steps: a) Initialization of one corresponding feature using vessel diameter information, b) optimization on three rotational and one translational parameter to register vessel structures that are represented as centerline graphs. We form a space of good features by iteratively creating new graphs from projected centerline images and by restricting the correspondence search only on branching points (the vertices) of the vessel tree. This algorithm shows good convergence and proves to be robust against deformation changes, which is demonstrated through studies on one phantom and three patients.

1 Introduction

Angiographic imaging is a widely used technique for intravascular interventions. In such treatments a preoperative 3D data set is usually acquired for diagnosis and planning. This data set shows detailed information of the patient's anatomy. 3D data sets are commonly acquired using CTA. During the intervention an intraoperative imaging device captures the current state of placed catheter and anatomy of the patient for navigation. In clinical practice, 2D fluoroscopic projections of the region of interest are acquired lacking spatial resolution compared to the preoperative data sets. Patients suffering from primary liver cancer are frequently treated with Transarterial Chemoembolizations (TACE). Here, in order to apply local chemotherapy and embolizing the blood vessels supporting the tumor, a catheter is inserted into the arterial vasculature in the hip region and guided to the tumor's location using DSAs. The navigation through the vessel system is rather difficult for physicians due to lack of depth perception and

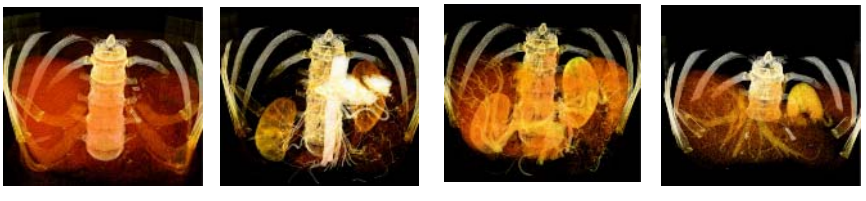
R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 873–881, 2006.

information about the tumor's location, which can only be visualized once the catheter is near the tumor and contrast injections propagate further down the vessel tree. Registering pre- and intraoperative data sets would allow physicians to view the catheter in 3D together with tumor location and detailed patient anatomy. Much attention has been drawn to the problem of 2D-3D registration of angiographic images. Intensity-based methods register two data sets by creating artificial X-ray projections (Digitally Reconstructed Radiographs, DRR) and optimizing cost functions directly evaluating pixel intensities [1, 2]. Feature-based approaches segment the vessels or extract vessel centerlines in order to use them as features for alignment optimizing distance-based cost functions. The former methods are very accurate but time-consuming due to expensive DRR generation (for our patient data, DRR generation takes approximately 1 second), while the latter ones are fast, robust, but often have to cope with segmentation errors and thus lack of accuracy. Since vessel structures are the dominant features in angiographic images, they are exclusively used for feature-based registration. Research is focusing on determination of corresponding points on vessel structures [3], choice of suitable metrics with (near) projective invariance [4, 5, 6, 7], and derivation of globally converging optimization procedures [8]. Hybrid methods register segmentations of (reconstructed) vasculature using intensity-based methods [9, 10]. All of these works address vessels in neurovascular images where rigidity is preserved or use gating for deformation compensation. To our knowledge, gating-absent 2D-3D registration in deformable regions like organs in the abdomen has not been performed yet. Moreover, the inherent tree property of vessel structures has not been used by these approaches. 2D-3D angiographic registration for liver interventions is an unsolved problem mainly because of two reasons. First, a suitable protocol for CTA to contrast liver arteries is not used in clinical practice. Second, an adequate registration algorithm which addresses the difficult task of aligning deformed vessel structures has not been developed yet. We present a novel one-click method for automatic rigid 2D-3D registration of angiographic images. We introduce a new CT protocol to visualize also liver arteries, which is the only visible structure during the interventional procedure. This protocol enhances the diagnostic value of CTA scans with interventional benefit and introduces a new way of intraoperative visualization. In a first step of the algorithm, 2D and 3D vasculatures are segmented using a combination of smoothing, background removal, and vessel enhancement filters followed by region growing. For the registration, graphs are created from both segmented data sets via a wave propagation algorithm [11] detecting vessel bifurcations as vertices and segments between junctions as edges. The algorithm then initializes one bifurcation correspondence via a-priori information of the patient location and near projective invariance of the vessel diameter. A nonlinear optimization is performed on a reduced parameter space iterating over 3 rotational and 1 translational parameter. Good feature correspondences are iteratively created for the matching by recurrent graph creation of the projected 3D centerline image. As will be shown, the registration algorithm is robust to deformation changes by focusing on the vasculature's bifurcations. Monte Carlo simulations on rigid phantom data and three patient data sets show convergence and robustness of the new algorithm. Graph representations of vessel structures have been used in 3D-3D deformable registration of different hepatic vessel trees [12, 13]. These approaches do not have to cope with projective invariance and can

use angles and lengths for their registration. Moreover, the hepatic vessel trees represent the portal venous system where the diameter of vessels and thus the number of bifurcations is higher compared to the arterial liver tree that is used in our approach.

2 CTA Protocol for Diagnosis and Intervention

The current workflow for TACE involves a three-phase CTA scan of the liver region consisting of native, arterial dominant, and portal-venous phase visualizing mainly the portal venous system, and the hepatic veins. However, during the intervention only arteries are visible in DSA, no portal or hepatic vessel systems, such that until now 2D-3D registration was not possible. We create a strong link between diagnostic radiology and intervention by defining a new protocol for the preoperative scans, introducing one additional run of an *angiographic* phase that visualizes liver arteries (delay times with bolus tracking: 6 (angiographic), 10 ± 2 (arterial dominant), 21 ± 4 (portal-venous) seconds), see fig. 1. The aim is to let the interventionalist benefit from 3D high resolution CTA scans for planning before, and for navigation during the intervention. We have acquired several data sets using a state-of-the-art 64-slice CT imaging device (Siemens Sensation 64). The spatial precision of the acquired scans could be reconstructed to $0.58 \times 0.58 \times 0.6 mm^3/voxel$ in a $512 \times 512 \times (280 - 500)\, voxel$ volume making it possible to extract vessels with a diameter down to $1 - 2mm$. As confirmed by physicians, the additional radiation exposure is acceptable for patients undergoing a TACE treatment. This new scan protocol is used to perform the registration. As a result, paths can be planned and intraoperative enhanced 3D navigation and visualization can be offered. To plan the path (roadmap) through the vessel system to the pathology on the new data set is one major issue for interventionalists. Registering the 3D pre- with the 2D intraoperative data allows to project this plan on the current DSA enabling physicians to discern vessel paths more easily.



(a) native (b) angiographic (c) art. dominant (d) portal-venous

Fig. 1. The new 4-phase CT protocol

3 Method

The presented method extracts a 3D and a 2D centerline graph from the vessel structures of the angiographic data sets and rigidly registers them using both, geometric and topological information. An initialization detects one point correspondence (the root). Then, an optimization is performed on a four dimensional parameter space combining projection, graph extraction, and cost function evaluation in each iteration.

3.1 Preprocessing

Segmentation. Since contrast agent is applied globally in a CTA scan, its propagation is uniform, but covers less of the vessel tree, opposite to DSA images where contrast is intraarterially applied. Moreover, the diameter of liver arteries is rather small (1-7 pixels), other than in 2D where magnification of the imaging device increases the size of tubular structures. Thus, we use two different methods in 3D and 2D for segmenting the vasculature. The 3D volume of the angiographic scan is smoothed using an anisotropic diffusion filter before the user has to place a seed point at the beginning of the arterial tree to be segmented. A region growing algorithm extracts the vasculature. In 2D, a background removal is performed by applying a bothat filter (closed image minus original image) followed by a multiscale vesselness filter assigning the probability of lying in a tubular structure to each pixel as described in [14]. A region growing algorithm extracts the vessel structure from a placed seed point.

Extraction of Graphs. From the segmented data sets one can easily extract a centerline image by applying thinning algorithms in 3D and 2D [15]. We extend a wave propagation algorithm [11] to cope with loops and use it for graph creation from this centerline. The graph $G = (V,E)$ is created with vertices at branching points of the vessel tree and edges in between (see fig. 2). Points between each bifurcation are sampled and stored as edge labels to keep edge position information. Since the thinning algorithm still introduces some wrong branches due to noise, edges smaller than the vessel diameter (which is also stored in G) are removed with a euclidean distance transform of the segmented data sets. For the same reason and in the same way very adjacent bifurcations are fused together.

3.2 Registration

Rigid 2D-3D registration aims at recovering the 6 degrees of freedom providing the viewing parameters of 2D image capture. The 6 DOFs are also referred to as the extrinsic parameters $[R|t]$ of a perspective projection, $x = PX = K[R|t]X$. $P \in \mathbb{R}^{3\times4}$ projects a homogeneous 3D point X onto a homogeneous 2D point x and can be decomposed into K (intrinsic parameters) and rotation R and translation t (extrinsic parameters). Since the interventional imaging device used (Siemens Axiom Artis) is fully calibrated, K is known. Moreover, distortion has already been compensated for inherently.

For given correspondences between 2D and 3D points ($x_i \leftrightarrow X_i,\ i = 1 \ldots n_1$), the solution to this problem is well-known: Find R,t that minimize the least-squares cost function

$$f_{lsq} = \sum_{i}^{n_1} ||x_i - \Phi_K(R,t,X_i)||^2,$$

where Φ_K is the projecting function with calibration matrix K. In our case, however, corresponding information is not available.

Initialization. State-of-the-art angiographic C-arm devices store lots of information concerning imaging geometry. Each 2D angiogram is provided with the calibration matrix K, a source-to-object (SOD)[1] distance, and an estimated magnification factor (s).

[1] The SOD is actually a source-to-table distance, but is called source-to-object distance in the DICOM header.

Moreover, it can be assumed that an image is acquired approximately from top-view visualizing the patient horizontally lying on the intervention table. With this a-priori information we can assume near projective invariance of centerlines and thus vessel diameters [6]. An initialization step tries to find a 2D correspondence to the bifurcation in 3D with the largest vessel attached, which is called the root. After setting the a-priori information we iteratively translate the 3D graph in x- and y-direction parallel to the image plane such that the root's projection is laid over 2D bifurcations attached to vessels with high diameter. The cost function f_{topo}, described further in this section, is evaluated and the (x,y)-translation yielding the lowest value is chosen. Thus, we find one corresponding vertex and reduce our optimization problem to recovering three rotational parameters (R) and one translational parameter t_{root} along the ray connecting X-ray source position to the projected root on the image plane.

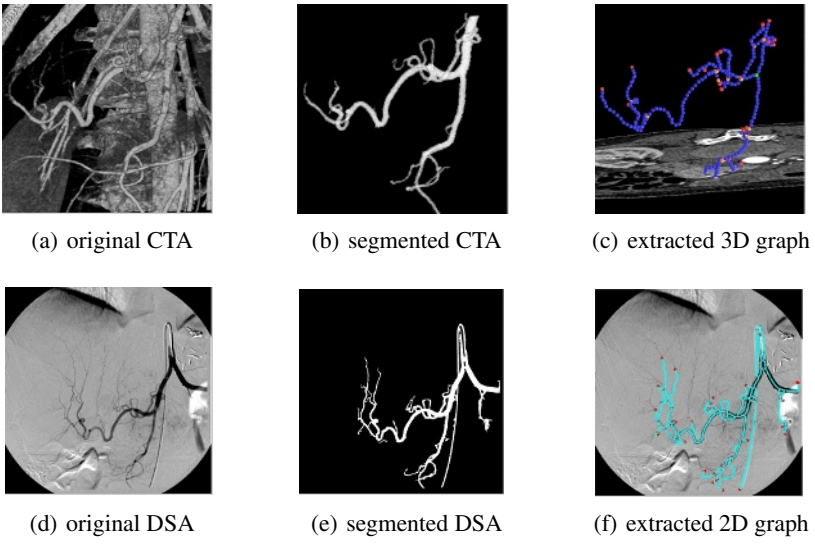
Geometric Registration. 2D-3D registration problems are twofold: recovery of corresponding information and estimation of rigid transformation. Since vessel structures can be interpreted as a number of curves, the well-known Iterative Closest Point (ICP) algorithm could be incorporated for finding both, matches and transformation parameters. Let $C(G,p)$ be a function determining the point of graph G closest to a given point p , the cost function is changed into

$$f_{icp} = \sum_{i}^{n_2} ||C(G_{2D}, \Phi_K(R, t_{root}, X_i)) - \Phi_K(R, t_{root}, X_i)||^2,$$

where $X_j,\ j = 1 \ldots n_2$ are all points representing the 3D vasculature (bifurcations and segment sampling points). Naturally, f_{icp} has many local minima since projected points of one 3D vessel segment could easily be driven to different, not corresponding 2D vessel segments in the optimization process. Moreover, even with outlier detection via an adaptive distance threshold based on statistical analysis, the cost function would yield wrong alignment due to deformation.

If we focus the correspondence search not on all curve points but only on dedicated ones that are likely to be detected in both data sets, we can improve the estimation process. Bifurcation points can be detected in 2D and 3D data sets very easily and represent good descriptors for a projection of vasculature since they are distributed over the whole vessel tree and are distinct to each other. Hence, we restrict our cost function only to these good features, the vertices $\{v_i\}$ of the graphs. Only inner vertices can be used since leaves in the graph account for the end of contrast propagation, which is different in 3D and 2D data sets. Unfortunately, segmentations and hence extracted graphs of the data sets are often very different due to local and global application of contrast agent. Moreover, the projection also produces crossings of vessels that cannot easily be detected and resolved since vessels can also be tangent to each other. This can be dealt with by performing a new graph extraction on the projected graph's centerline image in each iteration. The centerline image is created by drawing the 3D graph as a one-voxel-wide centerline in a volume and projecting the volume with the current imaging parameters.

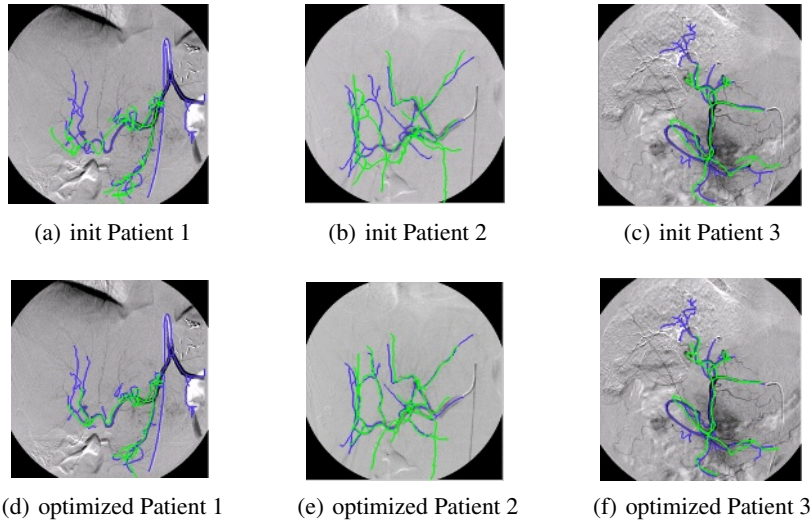
Topological Registration. Since we deal with graph structures, we can also incorporate topological information in the registration. Due to different segmentation results in 2D



(a) original CTA (b) segmented CTA (c) extracted 3D graph

(d) original DSA (e) segmented DSA (f) extracted 2D graph

Fig. 2. Preprocessing in 3D/2D. (a): volume rendered CTA, (b): segmented vasculature, (c): extracted graph; the green point is the root node, orange points inner, red points outer bifurcation points, blue points represent sampling points of the vessel segments. (d) and (e): original DSA and segmented vasculature.(f): 2D graph (turquoise are sampling, red are bifurcation points).



(a) init Patient 1 (b) init Patient 2 (c) init Patient 3

(d) optimized Patient 1 (e) optimized Patient 2 (f) optimized Patient 3

Fig. 3. Diameter initialized and registered pose of all patient data sets.

and 3D, the two graphs to be registered do not fulfill the subgraph property and it is not straightforward to do a topology-based graph matching to register the data. However, topological *tendencies* can be used for registration. The degree of each vertex can be

used as well as a time stamp coming from a breadth first search to penalize wrong bifurcation matches. In order to also involve the geometric meaning of the degree of each vertex, we determine the closest points to small edge segments (curves with up to 5 sampling nodes, we will call them *edgels*) attached to each bifurcation:

$$f_{topo} = \sum_{i}^{n} r_i^{bfs} \cdot r_i^{deg} \cdot ||C(v_j, \Psi_{K,i}(R, t_{root}, G_{3D})) - \Psi_{K,i}(R, t_{root}, G_{3D})||^2 + \sum_{t}^{m} ||C(G_{2D}^{edgel}, \Gamma_{K,t}(R, t_{root}, G_{3D})) - \Gamma_{K,t}(R, t_{root}, G_{3D})||^2,$$

where $\Psi = (v_1^{proj}, \ldots, v_{n_4}^{proj})$ projects the 3D graph G_{3D} with the current parameters R, t_{root}, and extracts a new graph from the projected graph's centerline image starting at the location of the projected root vertex. The resulting 2D graph's inner bifurcation list (without leaves) is returned by Ψ. r_i^{bfs} is the ratio of the normalized breadth first search values of the current bifurcation v_i^{proj} and the closest bifurcation in the 2D graph or its reciprocal if $r_i^{bfs} < 1$. r_i^{deg} is the ratio of the degrees of v_i^{proj} and the closest 2D bifurcation or its reciprocal if $r_i^{deg} < 1$. G_{2D}^{edgel} is a (disconnected) graph representing only bifurcations with edgels attached to them, and Γ returns the sampling nodes $(s_1, \ldots, s_m)$ of the edgels of the projected graph that has been created from the current centerline image. Since the cost function f_{topo} is highly non-linear a Downhill Simplex optimization was chosen.

4 Experiments and Results

Experiments have been carried out on a phantom head showing a rigid vessel structure and on three patient data sets. Ground truth registration has been created by intensity-based registration for the phantom head using Gradient Correlation as similarity measure and manually by physicians for the patient data. The initialization step was successfully performed for each of the data sets. The optimization procedure took 10-20 seconds for phantom and patient data.In order to evaluate the convergence and robustness of the optimization, Monte Carlo simulations have been carried out on all data sets. Random displacements in a range of $\pm 10^\circ$ in rotation angles and $\pm 10mm$ in translation along the axis connecting X-ray source position to root node have been added to the ground truth pose and the registration procedure was invoked. The following table shows standard deviations and root mean square errors (RMS) of 150 trials for each of the four parameters. Patient 1 has few deformation since the DSA was recorded at approximately the same breathing state whereas Patient 2 and 3 have rather large deformation. The larger deviation and RMS value for patient three is due to a more local distribution of the nodes in the DSA and CTA.

	$\sigma_\alpha[^\circ]$	$\sigma_\beta[^\circ]$	$\sigma_\gamma[^\circ]$	$\sigma_t[mm]$	$RMS_\alpha[^\circ]$	$RMS_\beta[^\circ]$	$RMS_\gamma[^\circ]$	$RMS_t[mm]$
Phantom	3.6	1.6	2.6	1.6	3.6	1.7	2.6	2.2
Patient 1	4.1	3.9	0.8	7.7	5.3	4.3	0.8	8.7
Patient 2	1.7	3.8	0.8	9.9	3.3	4.0	0.8	9.9
Patient 3	5.2	7.2	3.3	35.5	5.4	7.2	4.5	60.5

Fig. 3 shows the three patient data sets after initialization (a),(b),(c) and after optimization (d),(e),(f). The blue line shows the extracted 2D graph, the green line the projected 3D graph. As can be observed, the optimization procedure improves the initial alignment up to a deformation of the vessels. Thanks to the new CTA phase, interventionalists can plan the catheter insertion for TACE. At this point, the interventionalists are more interested in a precise rigid 2D-3D registration allowing them to execute their plan, and do not consider a deformable registration as vital. Therefore, the measure of success in this paper is a precise rigid 2D-3D registration up to a final deformation of vasculature

5 Conclusion

We introduce a new CTA scanning protocol for liver interventions resulting in a benefit for interventionalists in terms of depth perception and 3D planning to the diagnostic value of 3D high resolution images. By developing a 2D-3D registration algorithm based on the newly acquired data we enable physicians to transfer planning information to the interventional room. We overcome the difficulty of alignment by conceiving an algorithm, which iteratively generates good features that can be detected easily in 2D and 3D and describe the perspective projection adequately. The combination of geometrical properties of the vasculature like branching point coordinates or vessel diameter with topological properties of the vessel tree using an intuitive graph representation drives the registration process to an optimum quickly and robustly, if a good distribution of bifurcations over the image is provided. According to physicians the accuracy of the registration complies with the goal of plan transfer from 3D to 2D, i.e. from pre- to intraoperative data. Clinical evaluation and introduction of physically meaningful deformation will follow this work in the future.

Acknowledgements. This research was partially funded by an academic grant from Siemens Medical Solutions Angiography/X-Ray division, Forchheim, Germany. The authors would like to thank in particular Klaus Klingenbeck-Regn and Marcus Pfister for their continuous support.

References

1. Kerrien, E., Berger, M.O., Maurincomme, E., Launay, L., Vaillant, R., Picard, L.: Fully automatic 3d/2d subtracted angiography registration. In: MICCAI. (1999) 664–671
2. Turgeon, G.A., Lehmann, G., Guiraudon, G., Drangova, M., Holdsworth, D., Peters, T.: 2d-3d registration of coronary angiograms for cardiac procedure planning and guidance. Med. Phys. **32** (2005) 3737–3749
3. Kita, Y., Wilson, D., Noble, J.: Real-time registration of 3d cerebral vessels to x-ray angiograms. In: MICCAI. (1998) 1125–1133
4. Alperin, N., Levin, D.N., Pelizzari, C.A.: Retrospective registration of x-ray angiograms with mr images by using vessels as intrinsic landmarks. Journal of Magnetic Resonance Imaging **4** (1994) 139–144
5. Feldmar, J., Ayache, N., Betting, F.: 3d-2d projective registration of free-form curves and surfaces. In: Proc. Int'l Conf. of Computer Vision (ICCV). (1995) 549–556

6. Liu, A., Bullitt, E., Pizer, S.: 3d/2d registration via skeletal near projective invariance in tubular objects. In: MICCAI. (1998) 952–963
7. Chan, H., Chung, A., Yu, S., Wells, W.: 2d-3d vascular registration between digital subtraction angiographic (dsa) and magnetic resonance angiographic (mra) images. In: Proc. of the IEEE International Symposium on Biomedical Imaging. Volume 1205 of LNCS., IEEE (2004) 708–711
8. Florin, C., Williams, J., Khamene, A., Paragios, N.: Registration of 3d angiographic and x-ray images using sequential monte carlo sampling. In: Computer Vision for Biomedical Image Applications, First Int'l Workshop, CVBIA '05. Volume 3765 of LNCS., Springer (2005) 427–436
9. Hipwell, J., Penney, G., McLaughlin, R., Rhode, K., Summers, P., Cox, T., Byrne, J., Noble, J., Hawkes, D.: Intensity based 2d-3d registration of cerebral angiograms. IEEE Trans. on Medical Imaging (TMI) **22** (2003) 1417–1426
10. Vermandel, M., Betrouni, N., Palos, G., Gauvrit, J.Y., Vasseur, C., Rousseau, J.: Registration, matching, and data fusion in 2d/3d medical imaging: Application to dsa and mra. In: MICCAI. (2003) 778–785
11. Zahlten, C., Jürgens, H., Peitgen, H.O.: Reconstruction of branching blood vessels from ct-data. In: Eurographics Workshop of Visualization in Scientific Computing, Springer (1994) 161–168
12. Charnoz, A., Agnus, V., Malandain, G., Forest, C., Tajine, M., Soler, L.: Liver registration for the follow-up of hepatic tumors. In: MICCAI. (2005) 155–162
13. Aylward, S., Jomier, J.: Rigid and deformable vasculature-to-image registration: A hierarchical approach. In: MICCAI. (2004) 829–836
14. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale vessel enhancement filtering. In: MICCAI. (1998) 130–137
15. Palágyi, K., Sorantin, E., Balogh, E., Kuba, A., Halmai, C., Erdöhelyi, B., Hausegger, K.: A sequential 3d thinning algorithm and its medical applications. In: Proc. Int'l Conf. Information Processing in Medical Imaging (IPMI). Volume 2028 of LNCS., Springer (2001) 409–415

A New Registration/Visualization Paradigm for CT-Fluoroscopy Guided RF Liver Ablation

Ruxandra Micu[1,2], Tobias F. Jakobs[3], Martin Urschler[4], and Nassir Navab[1]

[1] Chair for Computer Aided Medical Procedures (CAMP), TU Munich, Germany
micu,navab@cs.tum.edu
[2] Siemens Medical Solutions, Computed Tomography, Forchheim
ruxandra.micu.ext@siemens.com
[3] Institute for Clinical Radiology, University of Munich, Grosshadern Hospital
tobias.jakobs@med.uni-muenchen.de
[4] Institute for Computer Graphics and Vision, Graz University of Technology
urschler@icg.tu-graz.ac.at

Abstract. 2D-3D slice-to-volume registration for abdominal organs like liver is difficult due to the breathing motion and tissue deformation. The purpose of our approach is to ease CT-fluoroscopy (CT-fluoro) based needle insertion for the Radiofrequency Liver Ablation procedure using high resolution contrasted preoperative data. In this case, low signal-to-noise ratio, absence of contrast and additional presence of needle in CT-fluoro makes it difficult to guarantee the solution of any deformable slice-to-volume registration algorithm. In this paper, we first propose a method for creating a set of ground truth (GT) simulation data based on a non-linear deformation of the CT-fluoro volume obtained from real patients. Second, we split the CT-fluoro image and apply intensity based rigid and affine registration to each section. We then propose a novel solution, which consists of intuitive visualization sequences of optimal sub-volumes of preinterventional data based on the registration results. Experiments on synthetic and real patient data and direct feedback of two interventionalists validate our alternative approach.

1 Introduction

Radiofrequency Ablation (RFA) of the liver is an interventional procedure applied to patients with unresectable primary liver tumors (HCC - hepatocellular carcinoma) and metastases. The percutaneous needle insertion is performed using different imaging guidance systems like CT, CT-fluoro, MRI or Ultrasound [1]. We are investigating the registration of the CT-fluoro slices with the preinterventional CT data-set (volume).

Related Work. Xu et al. [2] use 2D CT-fluoro to 3D CT volume intensity based registration for tracking pulmonary lesions. However, the tracking is based on a region rich in texture which cannot be applied in case of the liver. Xu et al. [3] estimate the deformation of CT slices of lung to CT volume by dividing the image into overlapping small disks. Note, that these lung applications did not use contrasted preinterventional data and have richer texture than liver.

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 882–890, 2006.

Combinations of features like vessels with intensity based methods for registering US slices to preoperative MR volumes also for RF liver ablation is reported in Penney et al. [4]. Unfortunately vessels are not visible in CT-fluoro slices in order to be used.

Major issues and main contribution. The complexity of this registration task is described below:

1. RFA is performed while the patient is breathing continuously, whereas the preinterventional CT volume is usually acquired during deep inspiration. One reason is because only in this breathing stage the pleura cavity can be appreciated to its full extension, and the other reason is that the patient can hold his breath for the imaging procedure more easily. Breathing instructions during the intervention are common. Once the needle is inserted into the target lesion breathing is usually impaired due to moderate pain and/or psychological reasons. A review regarding liver motion and deformation due to respiration was published in [5].
2. CT-fluoro slices have a lower signal-to-noise ratio compared to preinterventional CT data due to the lower radiation dose applied.
3. A single fluoro slice provides much less information than an entire CT volume. Futhermore it is thicker than preinterventional CT data.
4. The only anatomical features that could be used for liver registration are the vessels which are only seen in the contrast-enhanced preinterventional CT scan. Neither intrahepatic vessels nor metastases can be appreciated using CT-fluoro slices. Since primary liver cancer (HCC) is usually pretreated using transarterial chemoembolization (TACE), the used radiopaque embolic agent can be visualized in the non-enhanced CT-fluoro scan.
5. Registration needs to be done in real time.

In addition, no ground truth (GT) data are available. The generation of gold-standard data by incorporating fiducial markers [6] in the liver is, due to the minimal-invasive nature of this procedure, not feasible. This paper includes two main contributions. The first is the elaborate generation of a synthetic data set allowing precise analysis and testing of the registration solutions. For this purpose, approximate fluoro slice simulations were derived from preinterventional CT volume to evaluate registration algorithms. The creation of this synthetic data is described in section 3.1.

Two experienced interventional radiologists independently defined planes in the preinterventional CT volume that they considered to be the best approximation to the corresponding fluoro slices. A variation considering the inter-observer agreement in defining corresponding slices was clearly observed. Partner physicians believe that there is also high chance for intra-observer disagreement, in case one interventionalist is asked to redefine the slices at a later time. Based on this finding and the major issues described above, we believe that a 2D-3D non-rigid registration solution precisely deforming the high quality contrasted 3D volume based on the single low quality not contrasted CT-fluoro cannot be guaranteed at the level required for medical procedures.

The second contribution of this paper is therefore to introduce an alternative solution. Rigid and affine models do not capture the deformation [5]. However, we propose a novel solution, which consists of intuitive visualization sequences of optimal sub-volumes of preinterventional data based on rigid and affine registration results of splitted CT-fluoro slices. This sub-volume is automatically visualized next to the CT-fluoro and contains all the preinterventional information related to the fluoro slice. Details are presented in 2.2.

2 Methods

2.1 Registration

Similarity measure. Due to the noise, the presence of the needle in the fluoro slices and the presence of contrast in the 3D volume a robust intensity based similarity measure, f is chosen [7], where high differences in intensity are decreased. λ controls the sensitivity of this bell shaped function.

$$f = \sum_{i,j} \frac{\lambda^2}{\lambda^2 + (I_1(i,j) - I_2(i,j))^2}$$

Registration details. The initial pose in the volume is set at the axial position where the fluoro slices are extracted in the deformed and noisy volume. This is analogous to initialization of real fluoroscopic slices using the DICOM header 'Image Position Patient' since the patient remains on the same table for the preinterventional CT volume scan and the intervention. RMS measurements between initial pose in the volume and defined pose by physicians vary between $26mm - 38mm$, but this initial position proved to be best for convergence towards the target. A multiresolution approach is used to avoid local minima. The parameters for the rigid (6 DOF: 3 translations, 3 rotations) and affine transformation (12 DOF: rigid, 3 scalings, 3 shearings) of the CT volume, estimated at each resolution are used as initial values for the higher resolution. As optimization schemes we used the Best Neighbor and Powell-Brent. Real fluoro slices are additionally filtered with a Gaussian filter to reduce the noise and the artefacts of the needle.

2.2 Volume Visualization Based on Piecewise Rigid and Affine Registration

Dividing the liver into two parts, the upper part containing the ventral part of the liver and lower containing its dorsal part, enabled better results of the rigid/affine approximation. This is expected since the two parts undergo different level of movement and deformation. During expiration the rib cage volume decreases, thus increasing the pressure on the ventral aspect of the liver. The cut dividing the liver into upper and lower part is still manually positioned in the fluoro slice where the ligamentum falciforme is seen. The position of the cut is kept constant for all slices of one dataset. Sometimes during the intervention, the

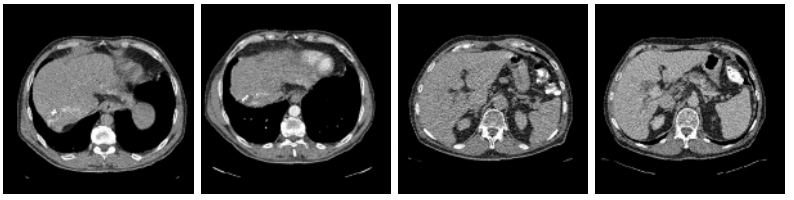
table is moved down allowing for more space for needle insertion. In this case it is needed to redefine the cut. Thus we are dividing the liver by a horizontal line into upper and lower part. These two parts are rigidly and affinely piecewise registered. In both models, rigid and affine, the estimated parameters by this piecewise registration define now the upper and lower planes. These resulting two planes should include a minimum volume with maximum information to be displayed in the operating room. The intersection line and angle between these two planes are calculated and then the target volume is shown as an animation by rotating 0.5° at each visualization step around the intersection line, starting from the upper plane to the lower and back. Up to now the interventionalists move out of the operating room to the workstation outside to scroll through the high resolution volume in order to orientate better. Providing an additional display in the operating room with this small target volume visualization would be of great value.

3 Results

3.1 Experiments with Simulated Fluoro Slices

Simulation strategy The simulation strategy consists of adding realistic noise as if scanning with low dose [8] and applying a simulated breathing deformation as proposed in [9] to the pre-interventional datasets. We are using two routinely contrast-enhanced datasets from a 4 detector-row CT, Siemens Somatom Sensation4. Since they were acquired in deep inspiration breathhold, we apply the forces of deformation model such way that we simulate another volume in expiration stage, the rib cage contracting, while the liver is moving towards the heart. The two datasets were deformed with a translational force, representing the diaphragm movement, of $30mm$ and with a radial force, representing the rib cage movement, of $5mm$. Six axial slices are extracted from each noisy deformed volume, from all parts of the liver with $1cm$ spacing in between. Fig. 1 shows one example from each dataset of real and simulated fluoro slices.

Noise addition For one dataset(dataset1) raw data are provided. This raw data is reconstructed again with Siemens Syngo IR Taskcard [8] twice. First time with parameters read from the DICOM header from the images reconstructed on the CT workstation in the hospital. Second time as if applying low dose to generate a noisy volume. Thus both volumes have the same field of view (FOV). For the second dataset (dataset2) raw data are not available. Therefore a software [10] is used to re-calculate the projections (raw data) of CT data. Noise is added to this raw data-set which is again reconstructed as noisy volume. Both datasets were downsampled in x, y from 512×512 to 256×256 to fit fluoro slices, which have an image size of 256×256. The resulting volume sizes are for dataset1: $256 \times 256 \times 208$ voxels with a voxel size of $1.32 \times 1.32 \times 1.00mm$, and for dataset2: $256 \times 256 \times 69$ with voxel size of $1.45 \times 1.45 \times 3.00mm$. Both volumes have a CT slice thickness of $3mm$.



(a) real (b) sim. (c) real (d) sim.

Fig. 1. 2D fluoro slices and simulations. (a) (b) - dataset1, (c), (d) - dataset2

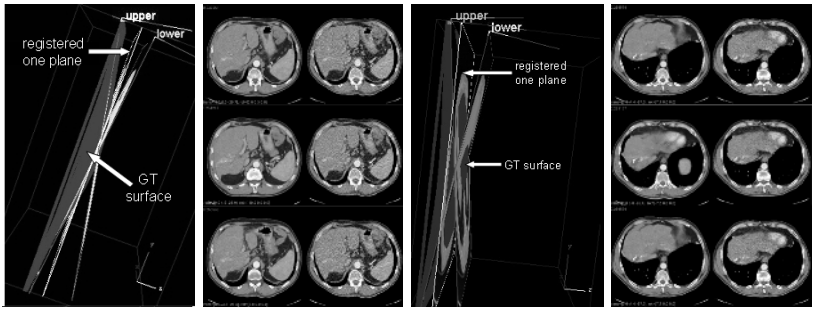
Table 1. RMS[mm] Dataset1 - $256 \times 256 \times 208$

	Best Neighbor						Powell-Brent					
	Rigid			Affine			Rigid			Affine		
Slice	one	lower	upper	one	lower	upper	one	lower	upper	one	lower	upper
1	12.98	5.90	4.35	8.27	5.59	4.95	10.51	7.68	5.90	6.75	9.17	5.42
2	11.08	5.77	4.61	9.86	5.83	4.95	10.36	6.96	6.43	6.95	6.24	7.39
3	9.61	5.84	4.76	10.52	5.84	5.02	12.56	6.72	4.46	6.86	5.81	5.09
4	9.69	5.95	4.90	9.72	6.17	4.89	9.19	7.21	8.17	7.57	6.25	8.36
5	9.59	6.19	10.59	9.60	6.22	5.10	10.58	5.80	4.30	6.90	7.51	6.19
6	8.20	6.53	4.73	9.51	6.43	4.85	8.53	6.37	7.75	8.07	5.88	5.36
Mean	10.19	5.84		9.58	5.49		10.29	6.48		7.18	6.56	

Registration details The registration starts at an initial resolution of $64 \times 64 \times 52$, increases to $128 \times 128 \times 104$ and finishes at $256 \times 256 \times 104$. Dataset2 is downsampled only in x and y direction due to the lower resolution along z. We are first registering the entire fluoro slice to the volume. The fluoro slice is mapped back with the known displacement field in the original volume. This curved, ground truth, surface is used to calculate the RMS error to the corresponding voxels estimated by the rigid and affine registration. Then each slice is divided into lower and upper part. Tables 1 and 2 show the results of one plane versus the results of upper and lower plane. The RMS error between the fluoro surface and the registered planes decreases by almost half except for Powell-Brent affine registration. Fig. 2 shows the rigid registration result planes of slice 2/dataset1 and 5/dataset1 with Best Neighbor optimization. The resulting planes of the registration are intersecting the ground truth fluoro surface. The upper and lower surface is better approximated by the 2 planes rather than by the one plane.

3.2 Experiments with Real Fluoro Data

Dataset2 is used here again. We selected 5 fluoro slices which were acquired during the intervention that contain needle, the tumor and different parts of the liver beeing imaged often during the intervention. For the 5 fluoro slices (256×256, $1.48 \times 1.48mm$, slice thickness $4mm$) we have independently defined



(a) 3D View (b) 2D Views (c) 3D View (d) 2D Views

Fig. 2. 3D view showing GT surface and registered upper, lower and one plane. 2D views showing from top to bottom the registered one plane, upper and lower.

Table 2. RMS[mm] Dataset2 - $256 \times 256 \times 69$

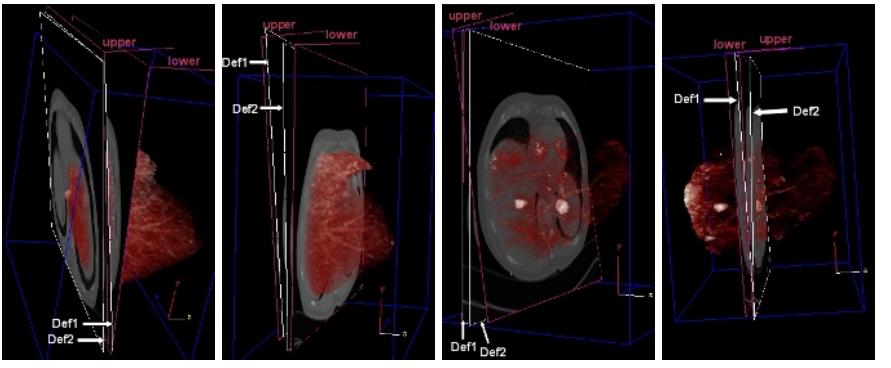
	Best Neighbor						Powell-Brent					
	Rigid			Affine			Rigid			Affine		
Slice	one	lower	upper	one	lower	upper	one	lower	upper	one	lower	upper
1	10.66	6.78	8.69	11.19	6.70	5.46	10.28	8.85	10.27	10.02	8.45	9.19
2	15.97	12.89	12.23	16.64	7.46	7.52	16.34	8.06	7.49	12.02	11.67	9.86
3	15.04	7.90	9.32	15.68	6.94	9.04	16.44	8.50	8.23	7.45	9.13	15.01
4	10.25	6.594	4.87	13.51	6.57	5.66	11.37	7.07	6.14	9.87	6.45	8.46
5	11.55	8.50	6.86	18.48	8.35	5.84	11.73	8.52	9.77	10.66	9.39	10.93
6	10.20	7.72	5.59	10.08	7.59	6.53	10.26	7.86	8.86	7.87	9.53	5.85
Mean	12.28	8.15		14.26	7.00		12.74	8.30		9.65	9.49	

with two physicians the planes that they consider the closest to the respective fluoro slices. The manual navigation included 6 DOF, 3 for translation and 3 for rotation. This dataset has a significant liver motion and deformation so it was hard to define the planes even by experts with manual registration. The other dataset which was used here, dataset3 ($256 \times 256 \times 158$, $1.29 \times 1.29 \times 1.5mm$) was acquired in shallow breathing. The liver motion is not so significant as in dataset2. We selected 3 fluoro slices (256×256, $1.29 \times 1.29mm$, slice thickness $4mm$): one with needle and tumor, one with tumor, one with the liver next to the heart. The variation between the defined slices by the physicians is shown below. Only liver voxels contribute to the RMS calculation.

	Dataset2					Dataset3		
Slice	1	2	3	4	5	1	2	3
RMS[mm]	17.16	10.25	5.79	16.50	19.68	4.12	7.05	9.08

The amount of voxels at each position $\mathbf{p}$ included in the volume of interest V defined by the upper plane: $\mathbf{n_u}, d_u$ and lower plane: $\mathbf{n_l}, d_l$, satisfy following inequality:

$$\left(\mathbf{p}\mathbf{n_l}^T - d_l\right)\left(\mathbf{p}\mathbf{n_u}^T - d_u\right) < 0$$



(a) ds2-slice3 (b) ds2-slice1 (c) ds3-slice1 (d) ds3-slice2

Fig. 3. Configurations of registered upper and lower plane, defined slices Def1, Def2 by the physicians and rendered liver surface for rigid registration with Powell-Brent optimization scheme. Percentage of intersection between liver voxels and the volume defined by the two physicians: (a) Def1 30%, Def2 0%, (b) Def1 100%, Def2 100%, (c) Def1 100%, Def2 79%, (d) Def1 79%, Def2 0%

Table 3. Dataset2 - % of included liver voxels for each defined slice in the sub-volume

	Percentage of Voxels included in VOI											
	Rigid Powell						Affine Powell					
Slice	1	2	3	4	5	Mean	1	2	3	4	5	Mean
%Def1	100	94.45	30.20	79.11	65.5	75.26	100	81.07	65.32	59.19	67.91	67.91
%Def2	100	96.09	0	74.52	11.81	60.01	96.12	87.87	2.82	61.17	14.13	55.73
	Rigid BestNeighbor						Affine BestNeighbor					
%Def1	72.55	73.57	44.00	71.00	78.81	68.95	88.14	24.44	5.58	68.89	65.50	54.00
%Def2	49.58	76.62	0	70.00	100	62.15	72.69	27.75	7.77	70.14	19.91	41.21

Table 4. Dataset3 - % of included liver voxels for each defined slice in the sub-volume

	Percentage of Voxels included in VOI							
	Rigid Powell				Affine Powell			
Slice	1	2	3	Mean	1	2	3	Mean
%Def1	99.60	78.93	100	91.37	98.88	100	45.63	79.96
%Def2	72.51	0	100	55.70	63.79	64.58	73.55	67.71
	Rigid BestNeighbor				Affine BestNeighbor			
%Def1	5.59	52.82	77.90	51.03	44.64	97.49	37.99	63.63
%Def2	69.45	0	83.16	48.66	69.3	6.22	53.09	39.77

Tables 3 and 4 show the percentage of liver voxels of the defined slices Def1 and Def2 by the two phyisicians, intersecting the visualized minimum volume. For real data the rigid model with Powell-Brent optimization works best. Particular slice 3 of dataset2 (ds2) poses difficulties for automatic registration due to a big rotation($\approx 12°$) about the x axis compared to the initial pose. But visual

inspection of the resulting registration and angle calculations of both defined slices to the closest registered plane of the minimum volume for this case (Def2: 0.77° and Def1: 1.96°), show that they lie very close to the volume. Fig. 3 (a) shows this configuration, (b) shows an ideal configuration as for slice 1. For slice 2 of dataset3 (ds3) we also calculated the angles of the defined slices to the registered closest plane. Additionally, the RMS error between liver voxels that are not included in the volume and the closest registered plane was calculated (Def1: 1.38°, RMS 1.92mm, Def2: 2.52°, RMS 5.65mm). Both RMS errors lie under RMS variation of Def1 and Def2 for this slice. This configuration is shown in fig. 3 (d).

4 Conclusion

In this paper we first presented major issues in providing a reliable 2D-3D registration between preinterventional contrasted CT and interventional non-contrasted CT-fluoro. We then presented a method for creation of GT data based on approximation of breathing and realistic noise models where we could evaluate our two part visualization-registration algorithm. This two part visualization-registration was then applied to registration with real CT-fluoro slices. In this way, we estimated the sub-volume of the preoperative CT data where pixels of the CT-fluoro image have originated from. This small size sub-volume of the high resolution preinterventional contrasted CT data is thus presented in the operating room as a short animation sequence next to the acquired fluoro slice. Results of the synthetic and real experiments and the feedback from two interventional radiologists confirmed the value of this alternative approach.

Acknowledgements. This research was funded by Siemens Medical Solutions, Germany within the VICORA project. The authors would like to thank in particular to Rainer Graumann, Michael Sühling (Siemens Medical Solutions) and Prof. Reiser (Institute for Clinical Radiology, Grosshadern) for their continuous support.

References

1. Friedman, M., Mikityansky, I., Kam, A., Libutti, S., Walther, M.M., Neeman, Z., Locklin, J.K., Wood, B.J.: Radiofrequency ablation of cancer. CardioVascular and Interventional Radiology (2004)
2. Xu, S., Fichtinger, G., Taylor, R.H., Cleary, K.R.: 3d motion tracking of pulmonary lesions using ct fluoroscopy images for robotically assisted lung biopsy. In: Medical Imaging 2004: Visualization, Image-Guided Procedures, and Display. Volume 5367., Galloway, R.L., Jr. (ed.) (2004) 394–402
3. Xu, S., Taylor, R., Fichtinger, G., Cleary, K.: Lung deformation estimation and four-dimensional ct lung reconstruction. In: Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention (MICCAI). Volume 3750 of Lecture Notes in Computer Science., Springer (2005) 312–319

4. Penney, G., Blackall, J., Hamady, M., Sabharwal, T., Adam, A., Hawkes, D.: Registration of freehand 3d ultrasound and magnetic resonance liver images. Medical Image Analysis **8** (2004) 81–91
5. Clifford, M., Banovac, F., Levy, E., Cleary, K.: Assessment of hepatic motion secondary to respiration for computer assisted interventions. Computer Aided Surgery **7** (2002) 291–299
6. Russakoff, D., Rohlfing, T., A. Ho, D.K., Shahidi, R., J.R. Adler, J., C. R. Maurer, J.: Evaluation of intensity-based 2d-3d spine image registration using clinical gold-standard data. In: Workshop on Biomedical Image Registration (WBIR). (2003)
7. Penney, G.P., Weese, J., Little, J.A., Desmedt, P., Hill, D.L.G., Hawkes, D.J.: A comparision of similarity measures for use in 2d-3d medical image registration. IEEE Transactions on Medical Imaging (TMI) **17** (1998) 586–595
8. Ulzheimer, S., Leidecker, C.: Syngo explorer image reconstruction(ir) taskcard, bericht zur validierung der option 'addition von rauschen'. VAMP Verfahren und Apparate der Medizinischen Physik **8** (2004)
9. Urschler, M., Bischof, H.: Assessing breathing motion by shape matching of lung and diaphragm surfaces. In: Medical Imaging 2005: Visualization, Image-Guided Procedures, and Display. SPIE, Galloway, R.L., Jr. (ed.) (2005) 440–452
10. Wessling, J., Fischbach, R., Esseling, R., Raupach, R., Heindel, W.: Effect of dose reduction and noise reduction filters on detection of liver lesions using multi-detector row ct. In: RSNA. (2005)

A New Method for CT to Fluoroscope Registration Based on Unscented Kalman Filter

Ren Hui Gong, A. James Stewart, and Purang Abolmaesumi*

School of Computing, Queen's University, Kingston, ON K7L 3N6, Canada
*purang@cs.queensu.ca

Abstract. We propose a new method for CT to fluoroscope registration which is very robust and has a wide capture range. The method relies on the Unscented Kalman Filter to search for an optimal registration solution and on modern commodity graphics cards for fast generation of digitally reconstructed radiographs. We extensively test our method using three different anatomical data sets and compare it with an implementation of the commonly used simplex-based method. The experimental results firmly support that, under the same testing conditions, our proposed technique outperforms the simplex-based method in capture range while providing comparable accuracy and computation time.

1 Introduction

Registration of CT to fluoroscope images is a fundamental task in Computer-Assisted Orthopaedic Surgery (CAOS) and Radiotherapy (CART). In the case of CAOS, registration of preoperative CT to a set of intraoperative fluoroscope images can be used to map the virtual patient (i.e. the preoperative CT) displayed on a screen to the physical patient in the operating room (OR). The registration precisely links the points in CT with the corresponding points in OR so that the CT can be used to guide the intervention. In the case of CART, registration of CT to portal images allows precise configuration of treatment beams so that the radiations are focused on tumors/lesions thus the damage to healthy tissues remains minimal.

The CT to fluoroscope registration problem can be briefly described as finding a geometric transform that brings CT, denoted *moving data*, to the patient's coordinate frame, in which fluoroscope images, denoted *fixed data*, are captured on the fly during the surgery or treatment. It involves determining the type of the transform as well as an initial guess of its parameters, defining a similarity measure between CT and fluoroscope images, and selecting an optimization method that updates the parameters iteratively until an optimal similarity value is achieved. A clinically usable CAOS or CART system requires accurate, fast and robust registration between the fixed and the moving data.

A number of methods have been proposed to address the CT to fluoroscope registration problem in the past decade. Based on what type of information is used in registration, these methods can be roughly classified into three categories [1]:

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 891–898, 2006.

feature-based, intensity-based and hybrid methods. Feature-based methods extract geometric features, such as landmarks, curves and surfaces, from both CT and fluoroscope images and use them in the subsequent registration. This class of methods is often more efficient because reduced amount of data are involved. However, the errors in feature extraction propagate to errors in registration. Intensity-based methods directly use all or partial intensity values in the original images and no feature extraction is required. This group of methods tends to be more accurate and robust, but at the same time, their computation costs are usually more expensive. The hybrid methods have both advantages and aim to provide accurate and fast registrations.

CT to fluoroscope registration is a 2D-3D registration problem, in which one key challenge is to find an appropriate way to compare 3D CT with 2D fluoroscope. One solution is to generate intermediate simulated fluoroscope images, called digitally reconstructed radiographs (DRRs), from CT and compare them with the corresponding real fluoroscope images. Alternatively, one can reconstruct a 3D volume from the fluoroscope images and compare it with the 3D CT [1]. Majority of current methods fall into the first category, and have focused on fast DRR generation methods [2, 3, 4, 5] or robust similarity measures [6, 7, 8, 4] between DRR and fluoroscope, and have relied on simplex or gradient descent [8] to search for an optimal registration. In this paper, we propose to use the Unscented Kalman Filter (UKF) as the optimization method. UKF has some features that could potentially benefit the CT to fluoroscope registration problem, e.g. no derivatives calculation is required, dealing with multiple observations simultaneously, estimating the variance along with the state, etc. To validate [9] our approach, we extensively test our method on various phantom data sets and compare the results with those of a simplex-based approach.

The remaining of this paper is organized as follows: We explain our method in details in Section 2; Section 3 describes the testing scenarios and presents the experimental results; and Section 4 concludes the paper.

2 Methods

In this section we discuss the major constituents of our method, i.e. the spatial transform, the similarity measure, the UKF-based optimization algorithm, and a hardware-based DRR generation technique.

2.1 Transform

For CT to fluoroscope registration, a rigid transform is a natural choice in order to bring CT points into the patient's coordinate frame. The transform has six parameters: three Euler angles in degrees for rotation and three scalars in millimeters for translation. This representation was chosen because 1) it has fewer parameters compared with the cases that rotation is represented using quarternion or angle-axis, and 2) the rotation and translation components have same or similar dynamic ranges. The number of parameters and the scales

between parameters are important factors that may have significant impact on the performance and robustness of a registration algorithm.

An initial guess of the transform parameters is required to start the registration. This is done by manually selecting a few landmarks from CT, e.g. 3-4 points on the bone surface, and identifying their correspondences in the fluoroscope images, then the initial value is computed by solving an absolute orientation problem using singular value decomposition [10]. The chosen landmarks should be visible in both CT and fluoroscope images. If they are carefully selected and identified, the computed initial transform can yield an initial mean target registration error (mTRE) within 3cm, which is often enough to start our registration method.

2.2 Similarity Measure

Similarity measure is a metric that judges how similar the two compared data sets are. It is a function of the transform parameters and is computed from either the intensities or the extracted features of the two data sets being registered. The function can be linear or nonliner, and can have a single value or multiple values. For some similarity measures, the derivatives w.r.t. transform parameters can be computed at low cost, while for some other similarity measures, direct calculation of derivatives is not possible.

A variety of similarity measures [6, 7, 8, 4] have been suggested for comparing DRR with fluoroscope. The commonly used ones are normalized correlation (NC), variance-weighted correlation (VWC), mutual information (MI), pattern intensity (PI), gradient difference (GD), and gradient correlation (GC). These are highly nonlinear functions and, except for MI [8], explicit calculation of derivatives is not possible. Each function is suitable for a different situation and has different computation cost. Thus our method was designed in a way that all current measures can be used.

Due to the loss of depth information during fluoroscope acquisition, one single fluoroscope is often not enough to reach an accurate registration, thus multiple fluoroscope images from different viewpoints are used in our approach, which means our similarity measure is multiple valued and each value corresponds to a $\langle$DRR, fluoroscope$\rangle$ pair. Most current methods for CT to fluoroscope registration combine the values during the optimization process, in which information is lost.

2.3 Optimization

As most current similarity measures for CT to fluoroscope registration are nonlinear, no direct derivatives and multi-valued, we adopted UKF [11] as the optimization method in our approach. UKF is a sequential least squares optimization technique employed for solving nonlinear systems. It estimates both the state and the state variance, and deals with multiple observations simultaneously during the estimation process. It requires no derivatives, Jacobian or Hessian calculations. Instead, UKF uses the Unscented Transform (UT) [11] to carefully select a minimal set of sample points to learn about the behaviors of a nonlinear system.

The sample points completely capture the true mean and variance of the state and, when propagating through the true nonlinear system, capture the posterior mean and variance accurately up to at least the second order Taylor series approximation.

We use UKF to estimate the transform parameters in a chosen similarity measure. The system models (state and observation) have the following forms:

$$\begin{aligned} x_i &= x_{i-1} + N(\mathbf{0}, s_x^2) \\ y_i &= SM(x_i) + N(\mathbf{0}, s_y^2) \end{aligned} \tag{1}$$

where x is the transform parameters to be estimated, SM is the nonlinear similarity measure between the DRRs and the corresponding fluoroscope images, σ_x and σ_y are the variances of the process and measurement noises intrinsic to the system. As multiple fluoroscope images are used in our method, y is a multi-dimensional measurement vector. The UKF algorithm recursively estimates the state and its variance. Figure 1 outlines the algorithm.

Input:
- Initial state and its variance
- A sequence of observations (constant in our problem)
- UKF parameters

Initialize UKF with the provided UKF parameters
Until a satisfying SM value is achieved, do
- Generate sample points from the current state and variance using UT
- Propagate the points through the state and observation models sequentially
- Compute the gain using the results from the previous step
- Update the state and variance using the computed gain and current observations

Fig. 1. The Unscented Kalman Filter (UKF) algorithm

The UKF parameters have great impact on the convergence speed, and were determined empirically by try and fail. The observations are the target value of the chosen similarity measure, which are constant and known in our problem, e.g. 1.0 for normalized correlation.

2.4 DRR Generation

We use an improved hardware-based technique, i.e. the Adaptive Slice Geometry Volume Rendering Algorithm [2], to speed up the DRR generation process. The algorithm improves the common view-aligned 3D texture-mapping based method by adaptively slicing the volume based on image content. Firstly, the CT is partitioned into a set of axis-aligned bounding boxes (AABBs) based on a user-defined transfer function that removes the empty voxels. Then the AABBs are sliced perpendicular to the viewing direction (i.e. the focal axis of the C-arm) in order back to front. Finally, the powerful OpenGL features of consumer-grade graphics cards, including 3D texture mapping, multi-texturing, and fragment programs, are employed to render and blend the slices into a final

DRR. We tested our implementation on an ATI Radeon X800 card with 256MB video memory by rendering CT volumes of size 512x512x256 into DRRs of size 473x473, and have achieved the speed of 20-50 frames per second, depending on the image content in CT.

3 Experiments and Results

We evaluate our method using images of three different phantoms: a small scaphoid bone, a large pelvis, and a long and thin femur, all with embedded fiducial markers for gold-standard validations. A general simplex-based method was implemented along with the UKF approach for the purpose of comparison. For all the experiments, normalized correlation was used as the similarity measure, and the same set of observation and measurement noise assumptions were made for all UKF experiments.

We first ran experiments using synthetic fluoroscope images, i.e. DRRs, in which case the gold-standard is known. For each bone, a set of DRRs from three orthogonal views were generated with the CT positioned at the origin of patient's coordinate frame. To find whether viewing angle of fluoroscope affect the method or not, a second set of DRRs was also generated for the scaphoid bone by rotating the CT about axis (1,1,1) for 45 degrees.

Next we did experiments using real fluoroscope images and used the embedded fiducials for validation. Four metal markers were attached to the femur surface and visible on both CT and fluoroscope images. We reconstructed 3D markers from the fluoroscope images by back-projection, which were then registered to the CT markers using a fiducial registration [10]. The registration result was used as the gold-standard.

For each experiment, the corresponding method ran 100 times with the initial transforms generated by adding random rotations ($\pm 12^\circ$) and translations ($\pm 12mm$) to the gold-standard. We then recorded the initial and final mTREs. The mTRE was computed as the average difference between the positions of a point set mapped by the evaluated transform and the gold-standard transform. 20 randomly selected points from the region bounding the bone were used when calculating the mTREs. Figure 2 shows the initial and final mTREs for each experiment. Obviously, the UKF-based method consistently worked well for the full range (0-20mm) of initial errors, while the simplex-based method worked well only for the range (0-5mm).

A registration is said failed if the final mTRE is larger than 2mm for synthetic fluoroscope experiments, or 3mm for real fluoroscope experiments. We define *accuracy* as the mean of the final mTREs of all successful registrations, and *capture range* as the initial mTRE that at least 95% of registrations succeed if they are within the range. Figure 3 shows the distributions of the final mTREs for selected experiments. It is obvious that the two methods achieved similar accuracies. Table 1 shows the capture range for each experiment. Again the UKF-based method obtained larger capture ranges for all experiments.

Because the time spent in DRR generation is dominant in both methods, we use the average number of DRR generations as the measure to evaluate the

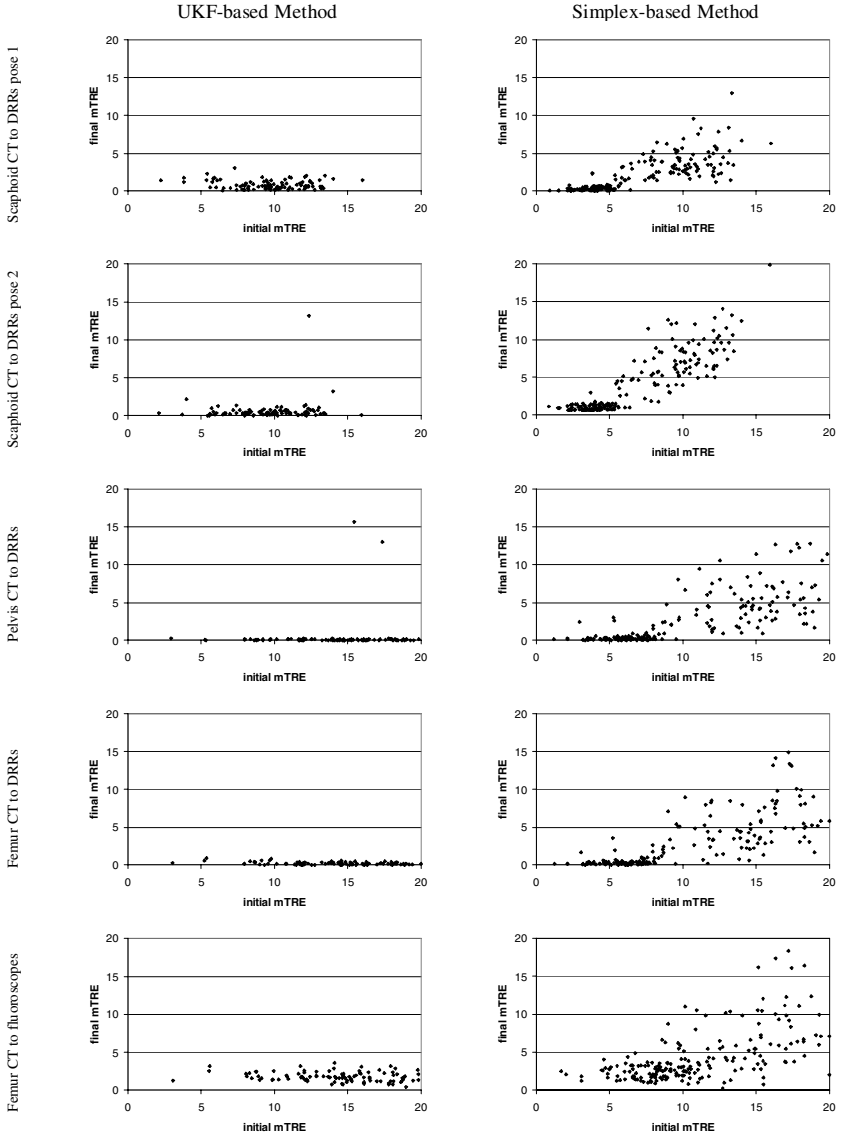


Fig. 2. Initial and final mTREs for each experiment. Left column: results of UKF-based method. Right column: results of simplex-based method. The first four rows show results for the CT to DRRs registration experiments, and the last row shows the results for the CT to fluoroscope registration experiments. All units are millimeters. Final mTREs above 20mm are removed from the graphs (approximately 3% of the total points for the UKF-based experiments of scaphoid and pelvis).

computation time. We have observed that, if initialized properly, the UKF-based method required slightly more DRR generations than the simplex-based method to reach a successful registration. But the differences were very small (within 150

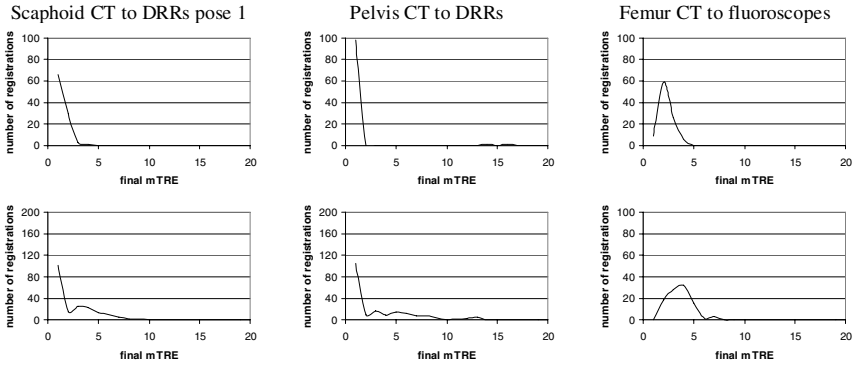


Fig. 3. Distributions of final mTREs for selected experiments. The first row shows the results for the UKF-based method, and the second row shows the results for the simplex-based method. The units of final mTREs are millimeters.

Table 1. Capture ranges for all experiments

	Scaphoid pose 1	*Scaphoid pose 2*	*Pelvis*	*Femur CT to DRRs*	*Femur CT to fluoros*
UKF	12mm	12mm	20mm	20mm	11mm
Simplex	5.8mm	5.5mm	9.6mm	9.2mm	8mm

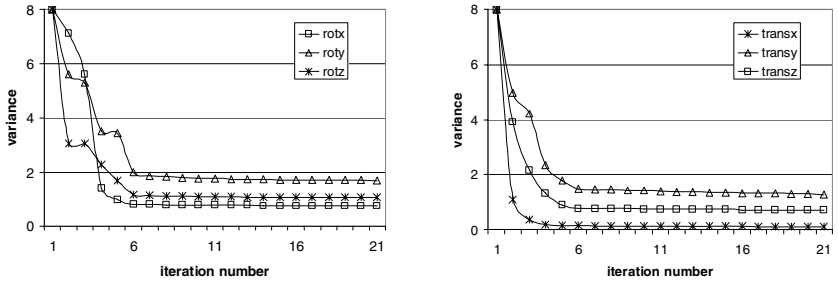


Fig. 4. The variance curves of one experiment (CT to DRRs registration, femur data). Left: the variance curves for rotation parameters. Right: the variance curves for translation parameters.

DRR generations, which takes less than 5 seconds for a graphics card producing DRRs at rate 30fps). We also observed that, if the initial position was far away from the gold-standard, at some point the UKF-based method required less DRR generations than the simplex-based method, which indicates that the UKF-based method has faster convergence speed.

The variance provided along with the state by UKF gives additional information about the confidence on transform parameters. If the variance of a parameter is small enough, we say that the estimated parameter value is accurate. The variance curves will tell whether the registration converges or diverges. Figure 4 shows

in one experiment how the variance of each parameter changes as the registration converges to the gold-standard.

4 Conclusions

We presented a new method for CT to fluoroscope registration that uses UKF to search for an optimal solution and a hardware-based volume rendering technique for fast DRR generation. The experimental results showed that our method outperformed the simplex-based method by having a larger capture range while providing comparable accuracy and computation time. In addition, the new method provides the variance of the resulting parameters as a by-product, which can be used as an additional metric to judge the confidence about the registration. We can conclude that, in the situations that simplex-based method could not perform consistently well, UKF is a preferable alternative choice for the 2D-3D registration problems.

Future work will include the extension of the method to register multi-fragment bone fractures simultaneously to a set of intra-operative fluoroscope images. Finding the UKF parameters in a systematic way is also an area deserving further research.

References

1. Tomazevic, D., Likar, B., Permus, F.: 3-d/2-d registration by integrating 2-d information in 3-d. IEEE Trans. Med. Imag. **25**(1) (2006) 17–27
2. Bethune, C., Stewart, J.: Adaptive slice geometry for hardware-assisted volume rendering. Journal of Graphics Tools **10**(1) (2005) 55–70
3. Russakoff, D., Rohlfing, T., Rueckert, D., Shahidi, R., Kim, D., Maurer, C.: Fast calculation of digitally reconstructed radiographs using light fields. In: Medical Imaging. (2003)
4. J.Bayouth, D.L., Kanade, T.: Transgraph: Interactive intensity-based 2-d/3-d registration of x-ray and ct data. SPIE Medical Imaging **3979** (2000) 385–396
5. Lacroute, P., Levoy, M.: Fast volume rendering using a shear-warp factorization of the viewing transformation. Computer Graphics **28** (1994) 451–458
6. Penney, G., Weese, J., Little, J., Desmedt, P., Hill, D., Hawkes, D.: A comparison of similarity measures for use in 2-d/3-d medical image registration. IEEE Trans. Med. Imag. **17**(4) (1998) 586–595
7. Livyatan, H., Yaniv, Z., Joskowicz, L.: Gradient-based 2-d/3-d rigid registration of fluoroscopic x-ray to ct. IEEE Trans. Med. Imag. **22**(11) (2003) 1395–1406
8. Zollei, L., Grimson, E., Norbash, A., Wells, W.: 2d-3d rigid registration of x-ray fluoroscopy and ct images using mutual information and sparsely sampled histogram estimators. CVPR **2** (2001) 696–703
9. van de Kraats, E., Penney, G., Tomazevic, D., van Walsum, T., Niessen, W.: Standardized evaluation of 2d-3d registration. In: MICCAI. (2004) 574–581
10. Tang, T., Ellis, R., Fichtinger, G.: Fiducial registration from a single x-ray image: A new technique for fluoroscopic guidance and radiotherapy. In: MICCAI. (2000) 502–511
11. Wan, E., van der Merwe, R.: The unscented kalman filter for nonlinear estimation. In: IEEE Symposium 2000 (AS-SPCC). (2000)

Automated 3D Freehand Ultrasound Calibration with Real-Time Accuracy Control

Thomas Kuiran Chen[1], Purang Abolmaesumi[1,2], Adrian D. Thurston[1], and Randy E. Ellis[1,3]

[1] School of Computing, Queen's University, Kingston, ON, Canada
[2] Dept. of Electrical and Computer Eng., Queen's University, Kingston, ON, Canada
[3] Dept. of Radiology, Brigham and Women's Hospital, Boston, MA, USA
purang@cs.queensu.ca

Abstract. 3D ultrasound (US) is an emerging new imaging technology that appeals to more and more applications in intraoperative guidance of computer-assisted surgery. In a freehand US imaging system, US probe calibration is typically required to construct a 3D image of the patient's anatomy from a set of 2D US images. Most of the current calibration techniques concern primarily with the precision and accuracy. However, for computer-assisted surgeries that may require a calibration task inside the operating room (OR), many other important aspects have to be considered besides accuracy. In this paper, we propose a novel system for automated calibration that is optimized for the OR usage with real-time feedback and control of the calibration accuracy. We have also designed a novel N-wire phantom, with greatly reduced complexity to facilitate mass production without compromising the accuracy and robustness.

1 Introduction

3D US imaging has seen increasing applications in intraoperative guidance of computer-assisted surgery. Compared to other imaging modalities (e.g., fluoroscopy, computed tomography (CT) or magnetic resonance (MR)), medical US is non-ionizing, compact and portable, relatively inexpensive, and capable of imaging in real time. To construct a high-resolution 3D US image of the patient's anatomy from a set of 2D images, a tracked 2D US probe that allows image acquisition in an unconstrained (freehand) motion is commonly used. Tracking is typically achieved by rigidly affixing the probe with a localizer traced by a position sensing system. However, knowing the position of the US probe alone is not adequate to determine the positions of the acquired 2D images. The relationship between these two coordinate frames can be calculated through a process known as US probe calibration by estimating a homogeneous transformation that maps the position of individual pixels from the US image frame to the US probe frame. The process is usually performed by imaging an artificial object with known geometries, referred to as a phantom. Calibration is therefore a fundamental step an a single point of failure in a freehand US system. A recent and comprehensive overview of the US calibration techniques could be found in [1].

Most of the current state-of-the-art calibration techniques focus primarily on precision and accuracy, which is well justified if the procedure is conducted in

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 899–906, 2006.

a laboratory setup; however, for computer-assisted surgeries that may require calibration inside the OR, many other important aspects have to be considered.

First, sterilization of the US probe needs to remove the spatial localizer. A re-calibration is therefore required in the OR, where a robust and fully automated procedure is always preferred. Overall, there exist two major types of calibration techniques: an iterative approach (e.g., wall, cross-wire and three-wire phantoms) vs. a closed-form solution (e.g., N-wire or Z-fiducial phantoms and the most recent, the Sandwich phantom [2]). Iterative approaches in general are less robust and less efficient than closed-formed solutions [1] and suffer from the non-guaranteed convergence, local minima and sensitivity to initial estimates. In addition, iterative methods typically need more images than closed-form solutions. For instance, the Cambridge phantom [3] (in the wall phantom group) requires capture of at least 550 images. In contrast, inspired by the work in [4], we have previously developed a custom-built N-wire phantom [5] to calculate the calibration parameters in a closed-form solution with less than 30 images and with sub-millimeter accuracy. Most recently, Boctor et al. [2] proposed a closed-form calibration method using a Sandwich phantom that requires as few as only three poses of US images. However, due to the generally poor visibility and the abundance of speckles in the US image, automated extraction of the wire points is always challenging for these landmark-based calibration techniques. For example, the current designs of the N-wire phantoms [4, 5, 6] typically require manual segmentation of the US images that is undesirable in the OR. Here we propose a robust and fully automated segmentation method to address this problem.

Another essential element that a conventional calibration system typically lacks for OR is the real-time feedback and control of accuracy. Calibration and validation are commonly two-phase tasks and remain isolated to each other in conventional techniques: first, a calibration is performed followed by a validation; hence, the only way for the surgeons to possibly improve an inaccurate calibration outcome reported by the validation procedure is to recalibrate, which is not only time-consuming but also, more critically, lacking assurance to warrant a satisfactory result in a repeated procedure. Boctor et al. [2] was among the first to address this issue in a *Bootstrapping* method where a closed-form calibration algorithm and a real-time validation procedure were close-looped to iteratively minimize the standard deviation of a 3D reconstruction error by reducing the errors and noise in the pre-collected US images. Reprocessing the pre-acquired data certainly has the advantage of limiting the number of input images to a minimum; on the other hand, depending on how much the image is retouched, it may also result in a final calibration matrix that is overfitting (or biased toward) the specific data collected for that particular calibration. In addition, since a small standard deviation of the 3D reconstruction error does not necessarily guarantee a satisfactory target registration error (TRE), using the error deviation to control calibration accuracy may not be sufficient enough. Inspired by their work, we propose an alternative real-time calibration-accuracy-control mechanism by iteratively minimizing the TRE directly while acquiring new US images for each iteration. This novel approach proves to bring a good balance between the number of input images and the calibration accuracy.

Furthermore, one of the major concerns in the OR is to sterilize the calibration phantom. Most of the current phantom designs are in general complicated and not optimized for such procedure. For instance, existing N-wire phantoms employ a large number of N-fiducials [4, 5]. To secure the wires, many holes are drilled on the phantom walls which provide potential hiding place for bacteria.

Finally, individual research groups undergo their custom designs of calibration phantoms that typically do not take costs into consideration. Manufacturing these phantoms for large-scale OR usage is not cost-effective and practical.

To the best of our knowledge, no current calibration system in the literature has ever attempted to address all these concerns and to provide a practical solution that is optimized to be used in the OR. In search for such a solution, we propose a complete system for automated freehand US calibration in the OR with real-time feedback and control of accuracy. We have also designed a novel N-wire phantom (*Kingston* phantom [7]) of greatly reduced complexity without compromising calibration accuracy and robustness. The phantom is cost-effective to make, with great potential for mass production and large-scale OR usage.

2 Methodology

Figure 1 gives a high-level system overview of our proposed US calibration solution. Besides *Kingston* phantom, the US calibration system consists of four essential components in an iterative closed-loop: *US image acquisition and tracking*, *automated segmentation*, *closed-form calibration*, and *real-time accuracy validation*. Serving as inputs, US images are continuously acquired from *Kingston* phantom and tracked in real time by a camera system. The positions of the N-fiducials in the US images are then extracted on-the-fly by an automated segmentation algorithm and together with their corresponding positions in the phantom space, they are fed into a least-squares-based closed-form calibration algorithm that we have previously developed [5] to calculate the calibration parameters. The calibration accuracy is then examined by the validation component in terms of TRE and fed back to the control loop to determine whether the new input data is acceptable and if the current calibration accuracy has reached a satisfactory level. When the TRE starts to converge after a certain number of input images and falls below a desired threshold, the system would terminate its iterations and output the final calibration result.

2.1 The *Kingston* Phantom

Kingston phantom [7] (Figure 2) consists of only a front and a back plate plus two side walls to form a simple cubic pipe. For effective sterilization, the phantom could be quickly disassembled and reassembled using an L-key screwdriver. To facilitate easy manufacturing, ensure rigidity and durability, and reduce distortion in the phantom geometry during the reassembling, the plates are in plain rectangle-shape and made of stainless steel at 5mm thickness. The front plate has an extended arm to mount a spatial localizer.

We have found, through experiments, that the speckles in US images which challenge automated segmentation are largely contributed by the destructive and

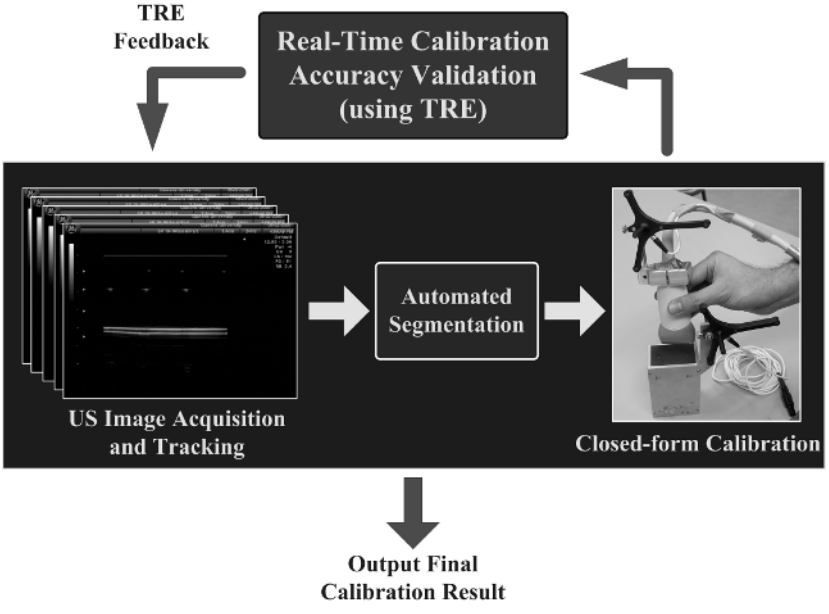


Fig. 1. The design of the proposed US calibration system

constructive interference of US signals traveling through too many N-fiducials arranged typically in parallel. We also discovered that the diameter of the nylon wire has non-negligible influence on the appearance of N-fiducials and at 0.35mm, it gives the smallest while the best defined dots in the image for the US probes we use. As a result (Figure 3), with the selected nylon wire, by reducing the number of N-fiducials to only two in double layers and shifting their positions apart, *Kingston* phantom produces considerably clean (less speckles) and very well defined images of N-fiducials that may facilitate the automated segmentation.

2.2 Robust and Fully Automated Segmentation

We have developed a fast, robust and fully automated segmentation algorithm to effectively extract the N-wire points in the US images acquired from *Kingston* phantom. The segmentation algorithm is specifically designed for *Kingston* phantom, however, the underlying principles could be easily generalized for all N-wire phantoms. The basic idea is to utilize the specific geometric appearance of the N-wires in the image, namely, the unique three collinear dots that form a typical N-wire intersection with the US image plane, and the two almost perfectly parallel lines that pass through these two layers of collinear dots. In general, the automated segmentation is composed of five major stages and employs various image processing techniques:

1. Grey-scale morphological operations are first used to remove large noisy parts, while retaining smaller objects in the image;
2. A second morphological pass is then employed to remove objects that are too small to represent true N-wire points;

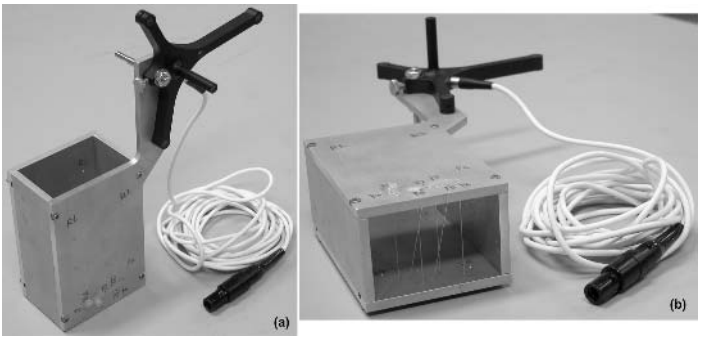


Fig. 2. Front (a) and side (b) view of *Kingston* phantom

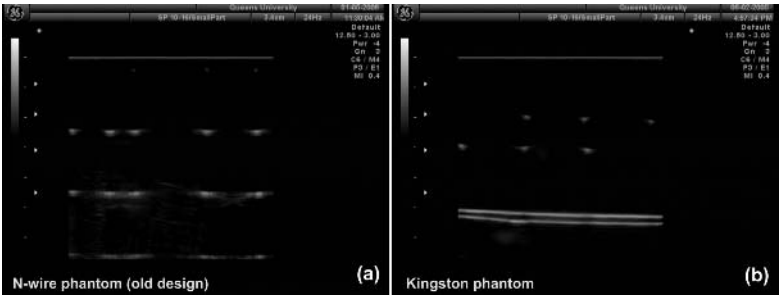


Fig. 3. US images of current N-wire phantom (a) and *Kingston* phantom (b)

3. The remaining pixels in the image are clustered into small regional spots with their centers representing a potential dot;
4. A line-search algorithm isolates all possible sets of three collinear dots from the others in the image;
5. Finally, a parallel-line-search algorithm locates the pair of parallel lines that are most likely represented by the N-wire points we are looking for.

To test the robustness of the segmentation algorithm to noise, 100 US images were acquired from *Kingston* phantom and we deliberately adjusted the settings of the US machine such that large areas of speckles were presented in the resulting images. The technique has proved itself to be very robust and accurate by correctly locating the true N-wire points in all 100 testing images. Figure 4 shows the segmentation results on one image that has worse-than-average visibility.

2.3 Real-Time Feedback and Control of Calibration Accuracy

Direct evaluation on calibration accuracy is challenging due to the lack of a reliable method to measure the exact spatial relationship of the US image frame to the probe frame. A common walk-around is to measure how close a 3D reconstructed position from the US image is (after applied the calibration parameters) to its actual position, frequently referred to as a Target Registration Error (TRE). Typically, stylus-probe measurement of the 3D position is treated as a

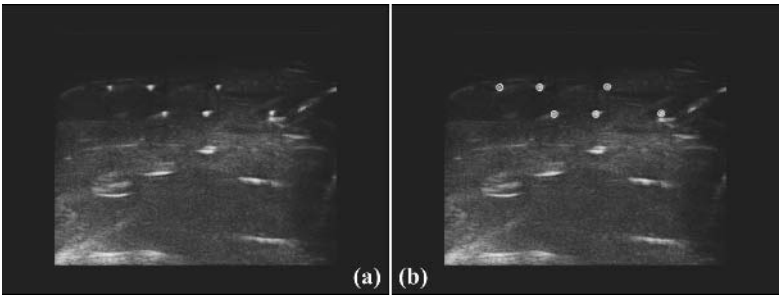


Fig. 4. Robust and automated segmentation results (b) of the original image (a)

golden standard and compared against the US measurement. However, it is impractical to obtain a TRE value for each calibration iteration using such manual procedures. Validation needs to be done automatically and in real time.

The uniqueness of a N-wire phantom is in its special geometry that allows us to identify some key physical features of the phantom both in the phantom space and on the US images, which essentially enables us to solve the calibration matrix in a single closed-form solution using least squares [5, 6]. The same principle could be used to automatically estimate the TRE for each iteration as in Eq (1):

$$\overline{TRE} = X_H - {}^{H}_{T}T \cdot {}^{T}_{P}T \cdot {}^{P}_{U}T \cdot X_U \tag{1}$$

where ${}^{P}_{U}T$ is the current calibration result to be validated, ${}^{H}_{T}T$ and ${}^{T}_{P}T$ are known by the spatial localizer affixed on *Kingston* phantom and the US probe, and X_H and X_U, the identified positions in the respective phantom and US image space. To avoid systematic errors, X_H and X_U are acquired separately for validation purposes only and should not be used as inputs to the calibration component.

3 Results

US images are generated by a General Electric (GE) Voluson 730 Expert US machine (Figure 5(a)). A Traxtal (Bellaire, TX) VersaTrax Active Tracker is mounted on a 2D linear array US probe (Figure 5(c)) and tracked by a Northern Digital (Waterloo, Ontario) Polaris Optical Tracking System (Figure 5(b)).

A total of 102 US images were acquired from *Kingston* phantom, then randomly selected as inputs to the calibration system. The number of input images ranged from a minimum of 4 to 102 (in which case all images were used). For every number of inputs, 100 independent calibration trials were conducted with images selected randomly per trial, and validated to obtain the TRE distribution. Figure 6 demonstrates the complete TRE distribution and convergence (for all 100 trials) with respect to the number of randomly selected input images along x-axis (the US signal propagating path), y-axis (inside the US image plane), and z-axis (along the US beam thickness direction).

The TREs are converged to 0.32mm, 0.13mm and 1.33mm in x, y and z axis, respectively, which is comparable to that of our old N-wire phantom [5]. The

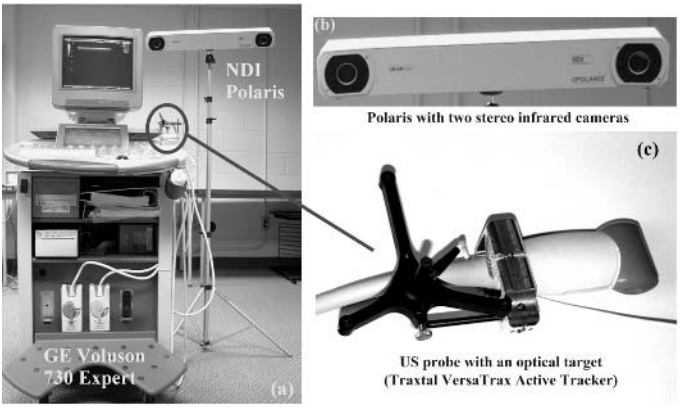


Fig. 5. Experimental setups

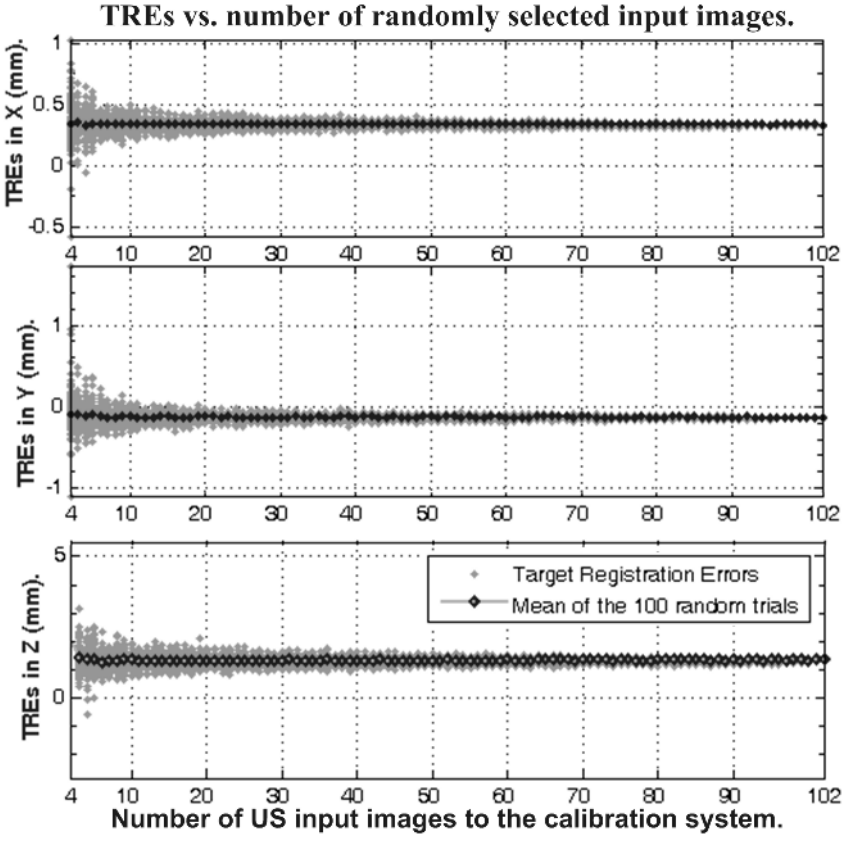


Fig. 6. Validation of the calibration accuracy

central black dots are the means of TREs for the 100 random trials at any given number of input images, which are essentially equivalent to the true converged TRE value. The gray dots are the actual TREs for each random trial and their outer band effectively outlines the converging patterns of TREs along each axis. As we can observe, the TREs start to converge quickly in both x and y directions to below 0.5mm only after 10 input images, and in z direction after 30 images to below 1.5mm. These results are as expected because z-axis is along the US beam thickness direction that has the most uncertainty among all three axes, therefore in general, it requires more images for the TREs to converge.

4 Discussion and Conclusion

We have presented a novel system for automated 3D freehand US calibration that is optimized for the OR usage with real-time feedback and control of calibration accuracy. The proposed system consists of:

- a novel *Kingston* phantom with greatly reduced complexity to assist OR usage without compromising the calibration accuracy,
- a robust and fully automated segmentation algorithm to extract wire points in the US images of the calibration phantom, and
- a real-time calibration-accuracy-control mechanism to iteratively minimize the TRE directly while acquiring new US image data at each iteration.

Our preliminary results have shown the system has a superior accuracy assurance balanced with real-time efficiency. The design of *Kingston* phantom is also cost-effective with the potential for mass production and large-scale OR usage.

References

1. Mercier, L., Langø, T., Lindseth, F., Collins, L.D.: A review of calibration techniques for freehand 3-D ultrasound systems. Ultrasound in Medicine and Biology **31**(2) (2005) 143–165
2. Boctor, E., Iordachita, I., Choti, M., Hager, G., Fichtinger, G.: Bootstrapped ultrasound calibration. Stud.Health Technol.Inform. **119** (2006) 61–66
3. Prager, R.W., Rohling, R.N., Gee, A.H., Berman, L.: Rapid calibration for 3-D freehand ultrasound. Ultrasound Med Biol **24**(6) (1998) 855–869
4. Pagoulatos, N., Haynor, D.R., Kim, Y.: A fast calibration method for 3-D tracking of ultrasound images using a spatial localizer. Ultrasound Med Biol **27**(9) (2001) 1219–1229
5. Chen, T.K., Abolmaesumi, P., Pichora, D.R., Ellis, R.E.: A system for ultrasound-guided computer-assisted orthopaedic surgery. Comput Aided Surg **10**(5) (2005) 281–292
6. Zhang, Y.: Direct surface extraction from 3D freehand ultrasound images. Master's thesis, The University of British Columbia (UBC), Vancouver, British Columbia, Canada (2002)
7. Chen, T.K., Abolmaesumi, P., Ellis, R.E.: An easy-to-use phantom with reduced complexity for automated freehand ultrasound calibration in the operating room and large-scale cost-effective manufacture. In Langlotz, F., Davies, B.L., Ellis, R.E., eds.: Proceedings of 6th Annual Meeting of the International Society for Computer Assisted Orthopaedic Surgery, Montréal, Canada (2006) 89–92

Non-rigid Registration of 3D Multi-channel Microscopy Images of Cell Nuclei

Siwei Yang[1], Daniela Köhler[2], Kathrin Teller[2], Thomas Cremer[2], Patricia Le Baccon[3], Edith Heard[3], Roland Eils[1], and Karl Rohr[1]

[1]Biomedical Computer Vision Group, Dept. Theoretical Bioinformatics, DKFZ Heidelberg, and University of Heidelberg, IPMB, Dept. Bioinformatics and Functional Genomics, Im Neuenheimer Feld 364, D-69120 Heidelberg, Germany
s.yang@dkfz.de
[2] Ludwig Maximilians University Munich, Biozentrum, Dept. of Biology II, Anthropology and Human Genetics, Germany
[3] CNRS UMR 218, Curie Institute, 26 rue d'Ulm, 75248 Paris Cedex 05, France

Abstract. We present an intensity-based non-rigid registration approach for normalizing 3D multi-channel microscopy images of cell nuclei. A main problem with cell nuclei images is that the intensity structure of different nuclei differs very much, thus an intensity-based registration scheme cannot be used directly. Instead, we first perform a segmentation of the images, smooth them by a Gaussian filter, and then apply an intensity-based algorithm. To improve the convergence rate of the algorithm, we propose an adaptive step length optimization scheme and also employ a multi-resolution scheme. Our approach has been successfully applied using 2D cell-like synthetic images, 3D phantom images as well as 3D multi-channel microscopy images representing different chromosome territories and gene regions (BACs). We also describe an extension of our approach which is applied for the registration of 3D+t (4D) image series of moving cell nuclei.

1 Introduction

For building maps of the 3D structure of chromatin fiber of the human genome in the cellular interphase, different cell nuclei need to be evaluated. The aim is to relate geometric information to genome function to obtain a better understanding of how chromatin structure affects gene regulation in normal as well as abnormal cells. Hence, one of the essential tasks is to analyze a large amount of multi-channel 3D microscopy images, which represent cell nuclei as well as genomic structures like chromosome territories and gene regions (BACs). However, since nuclei of different cells are used it is necessary to normalize these images, i.e., we have to put them into the same coordinate system. Normalization can be achieved by image registration and generally one has to cope with both global rigid transformations as well as local non-rigid deformations.

Non-rigid registration approaches can be divided in two main categories: geometry-based and intensity-based schemes. With geometry-based approaches

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 907–914, 2006.

first geometric structures such as points or surfaces are extracted from the images and based on that a non-rigid transformation is computed. One main difficulty is the finding of correspondences which often requires some kind of user-interaction. Intensity-based approaches, on the other hand, directly work on the image intensities and have the advantage that they are fully automatic. Recently, also hybrid approaches based on both geometry and intensity information have been proposed. The above-mentioned types of approaches have been widely applied in medical image analysis (see [1] for recent survey), however, application to biological image data is rare. According to our knowledge non-rigid registration schemes have not yet been used for 3D multi-channel microscopy images of different cell nuclei. An approach for aligning images of the same cell in different states has been described in [2]. For rigid registration of 3D cell images we refer to [3].

We have developed an intensity-based approach for non-rigid registration of 3D multi-channel microscopy images of different cell nuclei. Our approach is based on the demons algorithm [4], which has been successfully applied in medical image analysis. However, a main problem with 3D cell nuclei images is that the intensity structure of different nuclei differs very much, and thus an intensity-based scheme cannot be used directly. Therefore, in contrast to previous work, we first perform a segmentation of the images by global thresholding (after noise reduction using an anisotropic diffusion filter), smooth them by a Gaussian filter, and then apply an intensity-based algorithm. Since the computation time of the approach is relatively high, we use an adaptive step length optimization algorithm as well as a multi-resolution scheme which improves the speed significantly. Our approach is fully automatic and has been successfully applied using 2D cell-like synthetic images, 3D phantom images as well as 12 different 3D multi-channel microscopy images of cell nuclei. We also present an extension of our approach, which is applied for the registration of 3D temproal image series (4D datasets) of cell nuclei.

2 Non-rigid Registration of 3D Cell Nuclei

2.1 Preprocessing and Rigid Registration

To register 3D multi-channel microscopy images of cell nuclei we first reduce the noise in the images by applying a structure-preserving anisotropic diffusion filter [5]. Then we segment the 3D images by a global thresholding scheme, which is possible since the contrast in the images is relatively large. Afterwards we apply a rigid registration scheme. We have performed a systematic study of rigid registration algorithms for 2D and 3D images. We have tested different similarity metrics (mean squared error, normalized correlation coefficient, mutual information) and two different optimizers (regular gradient descent optimizer and Powell optimizer) [6] using 2D and 3D synthetic images, real images, and segmented images. We simulated translations and different rotations, and it turned out that the normalized correlation coefficient as similarity metric using the regular gradient descent optimizer yields the best results.

2.2 Non-rigid Registration of Different Cell Nuclei

For local shape alignment after global rigid registration we employ a variant of the demons algorithm [4], which is driven by symmetric forces. With $f(\mathbf{x})$ and $g(\mathbf{x})$ representing the intensity values in the source and target image at position $\mathbf{x}$, the instantaneous deformation vector field at iteration k can be written as:

$$d\mathbf{U}_k(\mathbf{x}) = \begin{cases} \frac{2[f(\mathbf{x}) - g(\mathbf{u}_{k-1}(\mathbf{x}))][\nabla f(\mathbf{x}) + \nabla g(\mathbf{u}_{k-1}(\mathbf{x}))]}{p(\mathbf{x})} & p(\mathbf{x}) \geq \epsilon \\ 0 & p(\mathbf{x}) < \epsilon \end{cases} \quad (1)$$

$$p(\mathbf{x}) = [\nabla f(\mathbf{x}) + \nabla g(\mathbf{u}_{k-1}(\mathbf{x}))]^2 + [f(\mathbf{x}) - g(\mathbf{u}_{k-1}(\mathbf{x}))]^2 \quad (2)$$

where ∇ denotes the nabla operator, ϵ is a small positive constant, and $\mathbf{u}_k(\mathbf{x}) = \mathbf{x} + \mathbf{U}_k(\mathbf{x})$ is the transformed position $\mathbf{x}$ with the total deformation field

$$\mathbf{U}_k(\mathbf{x}) = \mathbf{U}_{k-1}(\mathbf{x}) + d\mathbf{U}_k(\mathbf{x}), \qquad \mathbf{U}_0(\mathbf{x}) = \mathbf{0} \quad (3)$$

Our experimental results showed that this variant using symmetric forces yields better results compared to the standard approach, where the forces are not symmetric. Application of a prior rigid transformation improves the registration result and also reduces the overall computation time. Note, that the standard demons approach assumes that the intensities remain constant (optic flow based scheme), which is not true in our application since different cell nuclei significantly differ in substructure shape and intensity. Direct application therefore results in unrealistic compression and dilation. In contrast, we here consider cell nuclei as a whole and apply the algorithm to the segmented images. However, our experiments showed that we obtain good results for the contour regions of the objects but not for the inner (homogeneous) parts. To overcome this drawback, we suggest to smooth the segmented images first using a Gaussian filter. The purpose of smoothing consists in producing larger non-zero gradient regions in the neighbourhood of edges. Typically we use a relatively large value of the standard deviation of the Gaussian leading to strong smoothing ($\sigma = 6$). The deformation field obtained in this way is more homogeneous and appears to be more correct.

Adaptive step length optimization scheme. One problem still is the relatively high computation time of the algorithm, which is typical for intensity-based non-rigid registration schemes. For 3D images typically several hours are required. To accelarate the algorithm, we propose an adaptive step length optimization scheme which is described in the following. We note, that the demons algorithm can be interpreted as an iterative optimizing approach, where the step lengths during optimization are determined based on (1). However, in practise often the computed step lengths are too small and therefore many iterations are necessary to find the solution. We suggest to increase the step lengths and automatically adapt them during optimization. For adaptation we use a scheme which is similar to binary search [7] and we apply it for each voxel of an image during optimization. Binary search is a search technique for finding a particular

value in a list and it is an example of a divide-and-conquer algorithm. It is known that binary search is much faster than linear search.

While standard binary search is used to find a specific value in an order list of values, in our case we want to find the voxels in the target image whose intensities are most similiar to that of the considered voxels in the source image. Note that, in our application we can assume that there is only one object in an image (the cell nucleus) and for registration we use the segmented and Gaussian smoothed images. Therefore, the intensity values in the image are always increasing from the outside to the inner part of the segmented object, and thus are also sorted. Consequently, if we calculate the difference image between two images to be registered then this gives an indication of how good the chosen step lengths are. In our optimization algorithm we exploit this fact. If at a certain voxel there is a sign change of the difference image in subsequent iteration steps then the step length was chosen too large and the search exceeds the target position. At the beginning of the iteration the step lengths $\alpha_0(\mathbf{x})$ are determined based on the intensity difference between the two images $f(\mathbf{x})$ and $g(\mathbf{x})$ normalized by the maximal intensity value I_{max}:

$$\alpha_0(\mathbf{x}) = \frac{c[f(\mathbf{x}) - g(\mathbf{x})]}{I_{max}} \tag{4}$$

where c denotes a constant larger than 1 and determines the factor of the step lengths at the beginning of the iteration. For 8 bit images we have $I_{max} = 255$. Because of the normalization the values for the ratio in (4) (without considering c) lie between -1 and 1. It is obvious that the larger the geometric difference between two images, i.e., the further away the contours of the objects are, the larger can the step lengths be made. However, the step lengths should not be made too large, otherwise the algorithm well exceeds the searched target position. Based on our experience c should be smaller than 10, typically we used $c = 4$. During the iterative optimization process we automatically adapt the step lengths $\alpha_0(\mathbf{x})$. Whereas with standard binary search the search interval is halved in every iteration, in our case we only halve the step length when a certain condition is fullfilled:

$$\alpha_k(\mathbf{x}) = \begin{cases} \alpha_{k-1}(\mathbf{x}) & \alpha_{k-1}(\mathbf{x})[f(\mathbf{x}) - g(\mathbf{u}_{k-1}(\mathbf{x}))] \geq 0 \\ \dfrac{\alpha_{k-1}(\mathbf{x})}{2} & \alpha_{k-1}(\mathbf{x})[f(\mathbf{x}) - g(\mathbf{u}_{k-1}(\mathbf{x}))] < 0 \end{cases} \tag{5}$$

This condition has two terms. The first term is the previous (signed) step length $\alpha_{k-1}(\mathbf{x})$ and the second term is the intensity difference. If both terms have the same sign, i.e., if they are either both positive or both negative, then the search does not exceed the target position yet, and we can search in the same direction with the same step length as in the previous iteration. On the other hand, if the two terms have a different sign, then the search exceeds the target position. In this case we search in the opposite direction and halve the step length. In our approach the basic equation for image registration in (1) reads:

$$d\tilde{\mathbf{U}}_k(\mathbf{x}) = |\alpha_k(\mathbf{x})| d\mathbf{U}_k(\mathbf{x}) \tag{6}$$

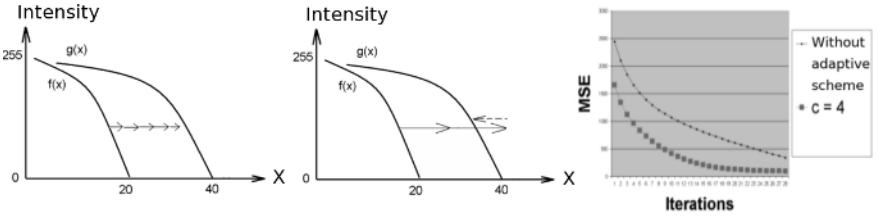


Fig. 1. Left: Optimization without adaptive step length scheme; Middle: Optimization using the adaptive step length scheme. Right: Comparison of convergence rates without and using the adaptive step length scheme ($c = 4$). MSE: mean squared error.

Note that we do not want that the algorithm is slower than the original one in case the intensity difference between two images is small. In this case $\alpha_k(\mathbf{x})$ is typically smaller than 1 and $d\tilde{\mathbf{U}}_k(\mathbf{x})$ would be smaller compared to the orignial formulation. Therefore, we impose the constraint that $\alpha_k(\mathbf{x})$ should always be larger or equal to 1. Thus, during iteration we use the condition $\alpha_k(\mathbf{x}) = 1, \quad \text{if} \quad \frac{|\alpha_{k-1}(\mathbf{x})|}{2} < 1$. Therefore, the α_k for all voxels will finally become 1, i.e., $\lim_{k\to\infty} \alpha_k(\mathbf{x}) = 1$, and (6) converges to the original form in (1). In summary, with our algorithm large step lengths are used if there are large shape differences. On the other hand, if there are small shape differences the original equation is used. Fig. 1 (left, middle) illustrates the use of the adaptive step length scheme for the 1D case. Without the adaptive step length scheme more than five iterations are required to align the source image $f(\mathbf{x})$ with the target image $g(\mathbf{x})$, whereas using the adaptive step length scheme needs only 3 iterations. In addition, Fig. 1 (right) demonstrates how the convergence rate is improved using the adaptive step length scheme ($c = 4$) in comparison to the standard approach (for the registration of 3D images of size $288 \times 288 \times 99$ voxels).

Multiresolution approach. To further reduce the computation time we have used a multi-resolution scheme, which follows the principle "from coarse to fine". Using this scheme we could reduce the computation time, e.g., for an image of size $512 \times 512 \times 24$ voxels from 225 min to 74 min without using the adaptive step length scheme and 35 min using it. For another image of size $256 \times 256 \times 33$ voxels we could reduce the computation time from 65 min to 15 min without using the adaptive step length scheme and 9 min using it. For running the algorithm we used a Linux Server with an AMD Opteron processor of 2.4 GHz and 7 GB memory.

For registering multi-channel 3D images we compute the transformation based on the first channel which includes the cell nuclei and then we apply the transformation to this channel as well as to the other channels which represent the different chromosome territories and gene regions (BACs).

2.3 Non-rigid Registration of Moving Cell Nuclei

The above described algorithm can be extended straightforwardly to be applicable to 3D+t (4D) multi-channel images of moving cell nuclei. The task here

is the normalization of the shape of moving cell nuclei over time in order to investigate the movement of certain protein particles within the cell nuclei. Given an image series, we first perform a rigid registration between consecutive images. For example, for time instances t to $t-1$ we determine the instantaneous rigid transformation matrix dR_t. The total rigid transformation matrix is then computed as the composition of previous matrices as

$$R_t = dR_1 \cdot dR_2 \cdot ... \cdot dR_{t-1} \tag{7}$$

Using this matrix we globally align the cell nucleus at each time point t w.r.t the first time point t_0. Second, we perform a non-rigid registration by determining the deformation fields for consecutive images, e.g., $\mathbf{U}_t(\mathbf{x})$ from time point t to $t-1$. The total deformation field is then calculated by composing these deformation fields. In addition, we regularize the composed deformation field by applying a Gaussian filter $G(\mathbf{x})$.

$$\mathbf{U}_t(\mathbf{x}) = d\mathbf{U}_{t-1}(\mathbf{x}) \circ \ ... \circ d\mathbf{U}_2(\mathbf{x}) \circ d\mathbf{U}_1(\mathbf{x}) \tag{8}$$

$$d\mathbf{U}_t(\mathbf{x}) \circ d\mathbf{U}_{t-1}(\mathbf{x}) \rightarrow G(\mathbf{x}) * [d\mathbf{U}_t(\mathbf{x}) + d\mathbf{U}_{t-1}(\mathbf{x})] \tag{9}$$

To speed up the calculations we use a Linux computer cluster and compute the transformations between two consecutive images in parallel.

3 Experimental Results

To systematically test our non-rigid registration approach we have generated 2D cell-like synthetic images. Application of the approach showed that symmetric forces lead to better results than asymmetric forces, because this variant takes into account the gradient information of both images instead of only that of one image. Next, we have validated our scheme using 3D CT images of the deformation of a silicon cube [8]. In the cube 343 beads were embedded and it was compressed vertically. The images of the cube before and after deformation (size of images: $288 \times 288 \times 99$ voxels) and the real displacements of the 343 beads are available. We have compared the displacements computed by our approach with the real displacements at these points, and obtained an acceptable correlation coefficient of $r = 0.96$ (for 12.5% strain).

Moreover, we have applied our approach to 3D multi-channel cell nucleus images representing different chromosome territories and gene regions (BACs). In total, we applied our approach to 12 multi-channel 3D images from two experiments. In the first experiment, we used 8 confocal 3D images from a 3D FISH (fluorescence in situ hybridisation) experiment with HeLa cells and using DAPI counterstain of the nucleus (image size $256 \times 256 \times 33$ voxels). We considered chromosome 4 and visualized the 3D registration result by overlaying the transformed segmented chromosome territories together with the surface of one of the nuclei as a reference. Fig. 2a illustrates that before registration some chromosomes (indicated by the arrows) lie outside the reference cell nucleus. After registration with our approach, all chromosomes are inside the reference

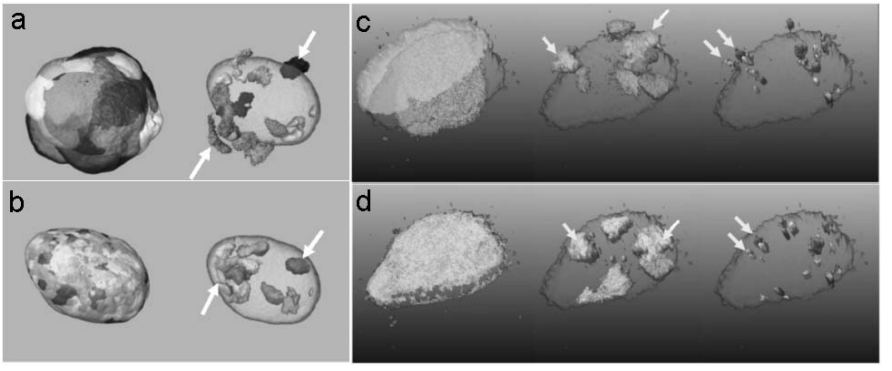


Fig. 2. a: 3D overlay of 8 different HeLa cell nuclei images (left) with chromosome 4 territories (right) before registration, **b:** Corresponding images after registration overlaid with a nucleus of one cell as a reference, **c:** 4 different HeLa cell nuclei images (left) with chromosome 2 territories (middle), and gene regions (right) before registration, **d:** Corresponding images after registration overlaid with a nucleus of one cell as a reference. The arrows point to corresponding chromosome territories und gene regions.

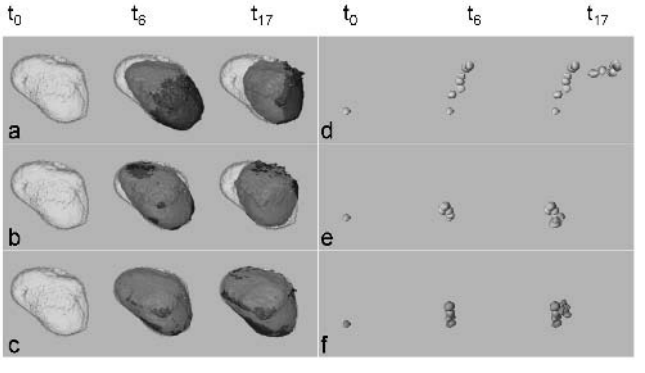


Fig. 3. 3D temporal image registration, selected images at time instances t_0, t_6, t_{17}, **a:** 3D view of one cell nucleus (white) as fixed reference nucleus (t_0), and moving nucleus over time (black), **b:** After rigid registration, **c:** After non-rigid registration, **d:** Positions of moving protein particle before registration, **e:** After rigid registration, **f:** After non-rigid registration.

cell nucleus and the chromosome structures from the different cells are located more closely together (Fig. 2b). In this way we achieve a normalization of the image data. In the second experiment with 4 confocal 3D images (image size $512 \times 512 \times 24$ voxels), where chromosome 2 territories as well as gene regions (BACs) were stained, we obtained the results in Fig. 2c,d. Again we achieve a normalization of the image data.

We have also applied our approach in a different application context, namely the registration of temporal 3D+t (4D) images series. We have applied our approach to three different images series and Fig. 3 shows the result for one 3D image series consisting of 17 images of size $256 \times 256 \times 75$ voxels. In Fig. 3a,b,c

are shown the original images, the result after rigid registration, and the result after non-rigid registration. In Fig. 3d,e,f are the corresponding positions of the protein particle. It can be seen that after registration there is much less movement of the particle since the global movement of the cell nucleus has been decoupled from the movement of the particle.

4 Summary and Discussion

We have presented an intensity-based non-rigid registration approach for the normalization of 3D multi-channel microscopy images of cell nuclei. The approach is based on a variant of the demons algorithm, uses segmented images, and employs a multi-resolution scheme. Furthermore, to improve the convergence rate we proposed an adaptive step length optimization scheme. We have successfully applied our approach using confocal multi-channel 3D images from different FISH experiments with HeLa cells. Moreover, a straightforward extension of our approach has been successfully used to register 3D+t (4D) images of moving cell nuclei.

Acknowledgement

This work has been supported by the EU project 3DGENOME. The work benefited from the use of the Insight Toolkit (ITK) [6]. We also acknowledge the support of the Human Frontier Science Program (HFSP).

References

1. Zitova, B., Flusser, J.: Image registration methods: a survey. Image and Vision Computing **21** (2003) 977–1000
2. Mattes, J., Fieres, J., Beaudouin, J., Gerlich, D., Ellenberg, J., Eils, R.: New tools for visualization and quantification in dynamic processes: Application to the nuclear envelope dynamics during mitosis. In: Proc. MICCAI. (2001) 1323–1325
3. Rieger, B., Mollenaar, C., Dirks, R.W., van Vliet, L.J.: Alignment of the cell nucleus from labeled proteins only for 4d in vivo imaging. Microscopy Research and Technique **64** (2004) 142–150
4. Thirion, J.P.: Image matching as a diffusion process: an analogy with Maxwell's demons. Medical Image Analysis **2** (1998) 243–260
5. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Trans. on Pattern Analysis and Machine Intelligence **12** (1990) 629–639
6. Ibanez, L., Schroeder, W., Ng, L., Cates, J.: The ITK Software Guide. Kitware Inc. (2005)
7. Knuth, D.: The Art of Computer Programming, Sorting and Searching. The third edn. Volume 3. Addison-Wesley (1997)
8. Kerdok, A., Cotin, S., Ottensmeyer, M.: Truth cube: Establishing physical standards for soft tissue simulation. Medical Image Analysis **7** (2003) 283–291

Fast Deformable Registration of 3D-Ultrasound Data Using a Variational Approach

Darko Zikic[1], Wolfgang Wein[1], Ali Khamene[2],
Dirk-André Clevert[3], and Nassir Navab[1]

[1] Chair for Computer Aided Medical Procedures (CAMP)
Technische Universität München, Germany
{zikic, wein, navab}@cs.tum.edu
[2] Imaging and Visualization Dept., Siemens Corporate Research
Princeton, NJ, USA
ali.khamene@siemens.com
[3] Department of Clinical Radiology,
University Hospitals Munich-Grosshadern, Germany
dirk.clevert@med.uni-muenchen.de

Abstract. We present an intensity based deformable registration algorithm for 3D ultrasound data. The proposed method uses a variational approach and combines the characteristics of a multilevel algorithm and the properties of ultrasound data in order to provide a fast and accurate deformable registration method. In contrast to previously proposed approaches, we use no feature points and no interpolation technique, but compute a dense displacement field directly. We demonstrate that this approach, although it includes solving large PDE systems, reduces the computation time if implemented using efficient numerical techniques.

The performance of the algorithm is tested on multiple 3D US images of the liver. Validation is performed by simulations, similarity comparisons between original and deformed images, visual inspection of the displacement fields and visual assessment of the deformed images by physicians.

1 Introduction

Being a non-invasive and cost effective modality, 3D ultrasound (US) has certain advantages over other modalities such as CT, MR or PET imaging. However, as confirmed by physicians, particular clinical applications of 3D US such as follow-up studies can be improved if the deformations between subsequent examinations can be undone.

Furthermore, since ultrasound is a real-time technique, it can be used in image-guided interventions. In this context, precise deformable mapping of the pre- and intra-operative images would facilitate the navigation process.

To this end, fast and accurate deformable registration methods for 3D US are required. This is a more challenging task compared to modalities such as CT or MRI since speckle, occlusion and other artifacts are inherent to the imaging technique. Hence, the registration method has to be able to cope with these

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 915–923, 2006.

modality-specific problems. Recently, a number of papers addressed these difficulties by proposing deformable registration algorithms.

Foroughi and Abolmaesumi [1] present a feature based approach, using attribute vectors as a similarity measure. These vectors contain image intensities smoothed by a Gaussian filter and computed on multiple resolution scales. A set of leading points is computed in the reference and the template image according to the distinction of the attribute vectors. Within this set, corresponding points are found by search in a preset range which yields a sparse displacement field. In the final step, the dense displacement field is computed by propagating the displacement values at the leading points using Gaussian kernels. Being among the most recent work on deformable 3D ultrasound registration, this paper reports runtimes of 430 seconds for images of the size $185 \times 113 \times 199$.

In [2] Xiao et al. present a scheme, using block matching and local statistics for estimating the local deformation of the data. A Bayesian regularization method is used for constraining the deformation. In a final step the dense deformation field is computed using B-splines. Here the algorithm runtimes were about 360 seconds for $149 \times 154 \times 121$ images. However, the weaker performance of the hardware (as of 2001) has to be taken into account.

Krücker et al. [3] developed a two-stage, sub-volume based method. In the first stage, a global affine registration is computed. In the second stage the subvolumes are registered locally, while the connectivity of the whole volume is achieved by an interpolation using thin-plate splines. Different similarity measures are used, and the reported runtimes were below 300 seconds (again it should be considered that hardware available in 2001 was used).

All these methods have in common that the displacement is computed only at certain points of the volume by finding correspondences between selected features and that they use an interpolation technique to determine the global displacement field. Feature based approaches however have certain intrinsic drawbacks. First, not all of the available information is used for registration. Furthermore, the search for good features is a complex issue. And finally, the search range for finding correspondences must be set, which results in either lower confidence levels for the results or higher computation times.

In this work we overcome these difficulties by directly computing the dense displacement field. To this end, we employ the known variational approach for deformable registration. To the best of our knowledge, this is the first use of such an approach for 3D ultrasound data. Furthermore, we show that although this approach involves solving a large system of partial differential equations (PDE), an efficient numerical implementation can reduce the computation time compared to the previously proposed methods.

Deformable registration based on the variational minimization of a functional is a well-known technique. A number of different similarity measures and possible regularization terms was proposed in literature, e.g. [4,5]. Many useful references are also to be found in the Computer Vision community, where the 2D Optical Flow problem leads to the same equations [6,4]. Especially the efficient solution of

the resulting PDE by using a multigrid solver is described in [7]. An introduction to multigrid methods can be found in [8].

2 Method

In the following we briefly present a deformable registration method based on the variational approach. The method yields a dense displacement field as the solution to the deformable registration problem.

In order to allow for larger displacements and to speed up the algorithm, we employ a multi-resolution pyramid. The pyramid is computed by reducing the resolution by a factor of 2 along each dimension from level to level by applying a discrete approximation to the Gaussian filter. On every level of the pyramid, the variational registration is performed using the result of the next coarser level as initial guess. Besides the general and well known advantages of this approach like accounting for large displacements, it is in particular beneficial when applied to ultrasound images.

Because of the speckle inherent to ultrasound images, these reveal large structures better when they are smoothed. Hence performing a large number of iterations on the coarser levels seems reasonable. This intuition is supported by the work presented in [1], where the entries of the attribute vectors were image intensities smoothed by a Gaussian filter and computed on different resolution scales. This directly corresponds to solving the problem on the coarser resolutions. Especially the fact that the intensities from the finest level are included only after Gaussian filtering suggests that in our approach no iterations on the finest level have to be performed. Furthermore, the results in [6] demonstrate that the variational registration should always be performed on smoothed images. In the validation section we confirm the benefit resulting from using coarser resolution levels.

Since in this work we consider deformable registration, we assume that the input images are already rigidly registered. Otherwise, they would be deformed in order to account for rigid motion which would lead to incorrect deformations. The rigid registration of the images, computing the pure rotation and translation parameters was performed using cross-correlation as similarity measure.

We now briefly present the variational method used on every level of the Gaussian pyramid.

2.1 Variational Deformable Registration

We set $\Omega = [0,1]^3 \subset \mathbb{R}^3$ as the spatial domain considered, and define the reference $R : \Omega \to [0,1] \subset \mathbb{R}$ and template image $T : \Omega \to [0,1] \subset \mathbb{R}$.

The registration problem is to compute the deformation function $\varphi : \Omega \to \Omega$, $\varphi(x) = x + u(x)$, s.t. the reference $R(x)$ and the deformed template $T(\varphi(x))$ are possibly similar with respect to a certain similarity measure. Here $u : \Omega \to \Omega$ is the displacement field.

Since the registration problem is ill-posed we do not solve it directly but we minimize a regularized energy functional

$$J[u] = D[u] + \alpha S[u] \tag{1}$$

with $D[u]$ a dissimilarity measure and a regularizer (smoothing operator) $S[u]$ which penalizes unlikely deformations [5]. The influence of the regularizer is determined by the scalar α. The solution of $u = \arg\min_{\hat{u}} J[\hat{u}]$ is the displacement field that represents the solution of the registration problem. The regularization term enforces that u is a diffeomorphism. This is an important property for the displacement field, since this class of functions maintains the topology of the images.

We use the sum of square differences (SSD) as dissimilarity measure [5]:

$$D[R, T, u] = \int_{\Omega} (R(x) - T(x + u(x)))^2 dx \quad . \tag{2}$$

The regularizer is chosen such that strong variations in the displacement field are penalized, i.e.

$$S[u] = \int_{\Omega} |\nabla u_x|^2 + |\nabla u_y|^2 + |\nabla u_z|^2 dx \quad . \tag{3}$$

This is the well-known diffusion regularization term, compare e.g. [4,5]. We solve the minimization problem (1) by applying the calculus of variations. This yields a partial differential equation (Euler-Lagrange equation), the solution to which is a necessary condition for a stationary point of the functional. With the above selection of the dissimilarity and regularization term we get the following non-linear partial differential equation.

$$\alpha \Delta u = f(u) \quad \text{with} \quad f(u) = (R - T(\varphi))\, \nabla T(\varphi) \quad . \tag{4}$$

This equation is the well known Poisson equation. For the solution of the system we assume the displacement function to be zero at the boundary, i.e. we impose homogeneous Dirichlet boundary conditions. For solving this non-linear PDE we apply a modified fix-point iteration, where only the incremental updates are regularized, compare [5].

2.2 Implementation Details

For the actual solution, we discretize the PDE (4) by the finite difference method using central differences [5]. After this step, the resulting system of linear algebraic equations $Au = f$ remains to be solved in every step of the fix-point iteration. This system is sparse, yet the size of the problem is extremely large. The matrix A is N×N with $N = 3n$ and n the number of the voxels in the volume. For a realistic problem size of 256×256×256 we end up with a system of the size $N \approx 5 \cdot 10^7$. Hence an efficient solver is needed. For such large systems one of the most efficient and general solvers is the multigrid method, which

has the theoretical complexity of $O(N)$ [8]. Furthermore, the multigrid solvers are very well understood for the Poisson problem and were already applied to registration problems before [5].

For solving the linear algebraic equations $Au = f$ we apply the Full Multigrid cycle (FMG) [8]. Experiments show that the number of pre- and post-smoothing Gauss-Seidel iterations for FMG can be chosen as low as 1.

For the fast implementation of the algorithm, it is required that the volumes have the size $S_1 \times S_2 \times S_3$ with $S_i = 2^{n_i} + 1$, hence the images are padded if necessary. Furthermore, we use the Euler reference frame for the displacement field. This enables simple and fast warping of the images within the algorithm. This is important since it is the most time-consuming step of the algorithm besides the solving of the equation system. We also assume that the diffeomorphism condition is sufficiently fulfilled, which allows for computing the inverse of the actual displacement field. This results in computing the registration from the reference to the template image. The force term then changes to $-(R - T(u))\,\nabla R$ which has the advantage of having to compute the gradient only once for the whole algorithm instead of once in every fix-point iteration.

The alpha values are chosen differently for each resolution level, such that the performed regularization is weaker on the finer levels. However, the same values can be used for all data sets. Empirical tests let us select $\alpha = 0.005$ for the finest level and increase it by factors $2 - 4$ for the coarser levels.

The stopping criterion for the number of fix-point iterations per level can be set manually, or a threshold ϵ for the relative decrease of dissimilarity can be defined. If the threshold is set in a conservative way (ca. 2% for coarsest level and 0.5% for finest), the number of iterations resembles the manually selected values.

3 Results and Validation

The volumes have been acquired from a healthy liver using a Siemens Sonoline Antares ultrasound system. We used a C5F1 curved-array mechanically swept 3D/4D ultrasound transducer at 3.5Mhz. Several volumes with large field-of-view were recorded at deep and medium inhale, and different orientations. For the evaluation we used 5 pairs of data sets. The digital scanline data stored by ultrasound device was converted to rectilinear volumes of size 257^3. The voxel size for the data sets was approximately 0.87, 0.72 and 0.59 mm/voxel along the respective dimensions.

The implementation was done in C++, and the experiments were performed on an AMD Athlon PC with 2.2GHz and 2 GB RAM.

The runtimes of the algorithm range from 10 to 40 sec. (av. 27 sec.), depending on the similarity of the images and thus on the number of iterations at respective resolutions levels. The computation time for one iteration is 15 sec. for the 257^3 resolution, 2.0 sec. for 129^3, 0.25 sec. for 65^3 and 0.02 sec. for 33^3.

3.1 Simulation Studies

For this test we use a known displacement field u to deform a given real US image T in order to artificially compute the reference image R. The images R

and T are then used as input for the registration algorithm. This way, we are able to use the field u as ground truth, to which the result u' of the registration procedure can be compared.

For the generation of a displacement field u, we use two different alternatives u_G and u_R. For the first option u_G we use a Gaussian $G_{\sigma,s}$ for every dimension of the field. Here, $G_{\sigma,s}$ has the standard deviation σ, and is scaled such that the maximum value equals s. We find that this gives deformations similar to the ones observed in the data. More challenging, the second option for the field u_R is to use the result of a previous registration process.

The results of this experiment are summarized in following table, where μ and σ denote the respective mean and standard deviation. In the first four columns we give the information describing the ground truth u, while the next three columns describe the difference between the ground truth and the computed displacement field $u - u'$.

type of u	$max(u)$ [mm]	μ_u [mm]	σ_u [mm]	$max(u-u')$	$\mu_{u-u'}$	$\sigma_{u-u'}$
u_G: $G_{0.2,5}$	9.035	0.096	2.174	2.341	0.052	0.170
u_G: $G_{0.2,20}$	36.138	0.386	0.543	11.243	0.161	0.461
u_R,	20.091	2.344	3.180	7.926	0.497	0.574

For the evaluation of the results, the mean error and the standard deviation of the error are the most important values. We observe a clear decrease in the mean error. The small standard deviation about the error mean describes the quality of the solution. It states that errors much larger than the mean and especially the maximum error rarely occur.

Furthermore, this test allows us to measure the contribution of the single resolution levels to the computed field u'. The results support the decision to reduce the number of iterations on the finer levels since the contribution to the ground truth field u is negligible and this at a high computational cost. Since the presentation of all possible combinations would be a five-dimensional problem, we list only some of the measurements, which illustrate the contribution of iterations on respective levels. In the left-most column, the table lists the number of iterations per level with the coarsest level first and the finest last. The tests were performed using ground truth field $G_{0.2,20}$ from the table above.

Row No.	Iterations per Level	$max(u-u')$	$\mu_{u-u'}$	$\sigma_{u-u'}$
1	(20 30 30 10 5)	11.243	0.161	0.461
2	(20 30 30 10 0)	11.277	0.164	0.466
3	(20 30 30 0 0)	12.343	0.179	0.529
4	(20 30 0 0 0)	22.427	0.338	1.164
5	(0 0 40 0 0)	14.210	0.203	0.657
6	(20 30 40 10 0)	10.261	0.161	0.438

The rows 1-4 show some results where we continuously decrease the number of iterations on the finer levels. In row 5 we observe that comparable results can be achieved by performing more iterations only on the middle resolution level.

Row 6 finally demonstrates that a combination of lower levels can be superior in terms of accuracy to using levels on the finest resolution.

3.2 Similarity Measurements

For the 5 data sets an increase in the similarity between the reference and the template before and after registration was measured. Here, we do not only measure the gain in the mean sum of square differences, but also compute Normalized Cross-Correlation (NCC) and Mutual Information (MI). Simply evaluating the SSD might be biased, since this measure is used for the registration itself. Comparison to more general metrics like NCC and MI [4,3] confirms the validity of using the SSD as a similarity measure for the registration of ultrasound images. In order to measure the gain in the similarity s we give the ratio of the differences of the similarity before and after the registration to the original similarity: $r = (s_{after} - s_{before})/abs(s_{before})$. In order to perform the comparison, the dissimilarity SSD is transformed to a similarity measure. For the SSD, the values for r ranged between 32.82% and 48.47% with an average of 40.78%. The results for NCC were 5.99%-10.94% with average 8.85% and for MI 21.21%-39.14% with average 30.49%. Additionally, we observe a strong correlation between the gain in the SSD and in NCC and MI.

3.3 Visual Assessment

Since the deformable registration is an inherently ill-posed problem, the inspection and visual assessment of the deformed images by a trained physician is of extreme importance for the validation.

First, the gain in the similarity and thus the benefit of the method for the medical use was verified. Besides using plain difference images, we deploy different colors (red and blue) for the two images and display the differences by fading between the images. This has the advantage that not only the similarity of the images but also the deformation can be perceived to a certain extent.

Second, the structure of the deformed data set was inspected to assure that the method did preserve the anatomy of the data. This test was also positive. Furthermore it was confirmed that the speckle pattern in the original data sets was not destroyed by the deformation. This is important since the speckle carries important diagnostic information about the imaged tissue.

Last, we display the computed deformation fields in order to verify that no apparently wrong displacements are computed. Standard visualizations of the displacement by displaying sampled vector fields or the deformed grid fail in 3D. The display is either too complex for the validation or the sampling is very low. Thus we display one slice of the deformed image and visualize the corresponding displacement vectors. This way, the origin of the voxels in the plane is visualized. This can be done with a much higher sampling than if the whole 3D field were visualized at once. This approach has the further advantage that the displacement is visualized simultaneously with the image data. So not only the coherence of the field but also its relationship to the anatomy can be verified.

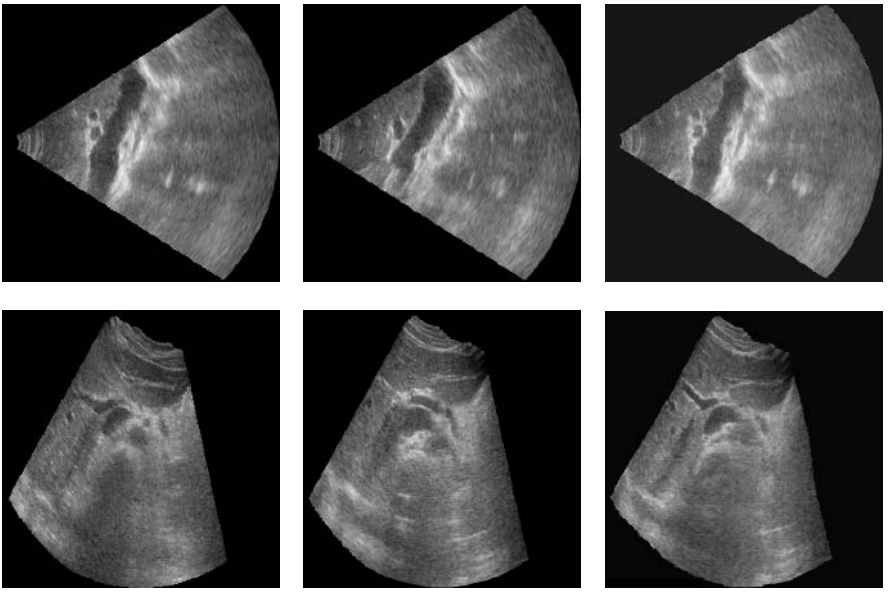


Fig. 1. Results of the deformable registration: the first column contains the reference, the second the template, and the third the deformed template image. The first row shows the results in the YZ and the second in the XZ plane. Corresponding slices are shown. It must be considered that apparently large deformations in the displayed slice are actually caused by smaller deformations along the axis orthogonal to the slice.

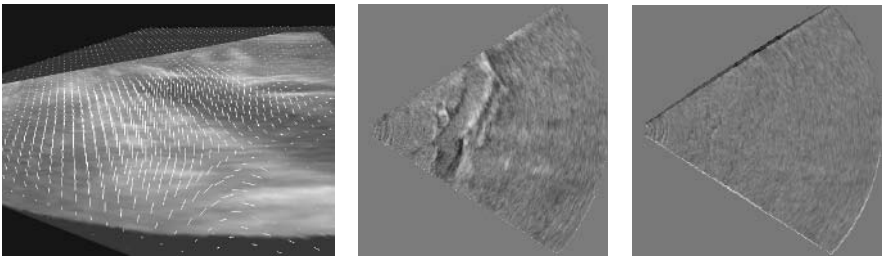


Fig. 2. Left: visualization of the 3D displacement field, any slice from the US volume can be selected. The two right images depict the difference images of the reference and template before and after registration.

Using the above tests, the significance of the proposed method as well as its validity were confirmed a physician.

4 Conclusion and Future Work

In this paper we presented a deformable registration method for ultrasound data. To the best of our knowledge this is the first time a variational approach is used

for 3D US data. It allows a direct computation of a dense displacement field even for large displacements and overcomes the difficulties inherent to feature-based methods. Besides, we demonstrate that using a Gaussian pyramid and efficient PDE multigrid solvers results in the currently fastest implementation. We evaluate the algorithm with a number of different methods and also present useful techniques for visual assessment of 3D deformable registration methods.

As future work we will evaluate the presented method on a large number of different data sets. The cooperation with medical partners will be extended in order to integrate the method into existing medical workflows or to define new possible applications. We plan to test the performance of other similarity measures and regularization operators within the variational framework on ultrasound data [4,5] and for inter-modality registration. Finally, we plan to tacle the specific US problem of dependence on scanning position. This can to different images depending on the scanning position, due to e.g. shadows from air or bones. We plan to integrate the results from [9] in order to solve that problem.

References

1. Foroughi, P., Abolmaesumi, P.: Elastic registration of 3d ultrasound images. In Duncan, J.S., Gerig, G., eds.: MICCAI. Volume 3749 of Lecture Notes in Computer Science., Springer (2005) 83–90
2. Xiao, G., Brady, M., Noble, J.A., Burcher, M., English, R.: Non-rigid registration in 3D free-hand ultrasound imaging of the breast. In Lemke, H.U., Vannier, M.W., Inamura, K., Farman, A.G., Doi, K., eds.: CARS. Volume 1230 of International Congress Series., Elsevier (2001) 1199–1200
3. Krucker, J.F., LeCarpentier, G.L., Fowlkes, J.B., Carson, P.L.: Rapid elastic image registration for 3-D ultrasound. IEEE Trans. Medical Imaging **21** (2002) 1384–1394
4. Hermosillo, G.: Variational Methods for Multimodal Image Matching. PhD thesis, Université de Nice - Sophia Antiplois (2002)
5. Modersitzki, J.: Numerical methods for image registration. Oxford University Press (2004)
6. Álvarez, L., Weickert, J., Sánchez, J.: Reliable estimation of dense optical flow fields with large displacements. International Journal of Computer Vision **39** (2000) 41–56
7. Bruhn, A., Weickert, J., Feddern, C., Kohlberger, T., Schnörr, C.: Variational optical flow computation in real time. IEEE Transactions on Image Processing **14** (2005) 608–615
8. Briggs, W.L., Henson, V.E., MacCormick, S.F.: A multigrid tutorial. 2. ed. edn. SIAM, Society for Industrial and Applied Mathematics (2000)
9. Wein, W., Roeper, B., Navab, N.: Automatic registration and fusion of ultrasound with ct for radiotherapy. In: MICCAI 2005 Proceedings. Lecture Notes in Computer Science, Springer (2005)

A Log-Euclidean Framework for Statistics on Diffeomorphisms

Vincent Arsigny[1], Olivier Commowick[1,2], Xavier Pennec[1], and Nicholas Ayache[1]

[1] INRIA Sophia - Epidaure Project, 2004 Route des Lucioles BP 93
06902 Sophia Antipolis Cedex, France
Vincent.Arsigny@Polytechnique.org
[2] DOSISoft S.A., 45 Avenue Carnot, 94 230 Cachan, France

Abstract. In this article, we focus on the computation of statistics of invertible geometrical deformations (i.e., diffeomorphisms), based on the generalization to this type of data of the notion of *principal logarithm*. Remarkably, this logarithm is a simple 3D vector field, and is well-defined for diffeomorphisms close enough to the identity. This allows to perform *vectorial* statistics on diffeomorphisms, while preserving the invertibility constraint, contrary to Euclidean statistics on displacement fields. We also present here two efficient algorithms to compute logarithms of diffeomorphisms and exponentials of vector fields, whose accuracy is studied on synthetic data. Finally, we apply these tools to compute the mean of a set of diffeomorphisms, in the context of a registration experiment between an atlas an a database of 9 T1 MR images of the human brain.

1 Introduction

In this article, we focus on the computation of statistics of general *diffeomorphisms*, i.e. of geometrical deformations (non-linear in general) which are both one-to-one and regular (as well as their inverse). To quantitatively compare non-linear registration algorithms, or in order to constrain them, computing statistics on *global* deformations would be very useful as was done in [6] with *local* statistics.

The computation of statistics is closely linked to the issue of the *parameterization* of diffeomorphisms. Many algorithms, as in [5, 9], provide dense transformations which are guaranteed to be diffeomorphic, and parameterize them via their displacement field. However, Euclidean means of displacement fields do not necessary yield invertible deformations, which makes Euclidean statistics on these parameters (or B-splines as in [10]) problematic for diffeomorphisms. In [8], it was proposed to parameterize arbitrary diffeomorphisms with Geodesic Interpolating Splines control points [4], and then to perform Euclidean operations on these low-dimensional parameters. However, although this guarantees the invertibility of the results, this may not be adequate for the whole variety of invertible transformations used in medical imaging.

To fully take into account the group structure of diffeomorphisms, it has been proposed to parameterize dense deformations with Hilbert spaces of *time-varying* speed vector fields, which yield geometrical deformations via the integration of

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 924–931, 2006.

an Ordinary Differential Equation (ODE) during one unit of time [14, 3]. In [15], it is suggested that the linear space of intial momenta of the geodesics of these spaces could provide an appropriate setting for statistics on diffeomorphisms. However, this is illustrated in [15] only in the case of landmark matching. To our knowledge, this statistical framework has not been used yet in the *general case*, certainly because of the *iterative nature* of the computation of the mean in this setting, which requires very stable numerical algorithms to converge.

In this article, we introduce a novel parameterization of diffeomorphisms, based on the generalization of the *principal logarithm* to non-linear geometrical deformations. Interestingly, this corresponds to parameterizing diffeomorphisms with *stationnary* speed vectors fields. As for matrices, this logarithm is well-defined only for transformations close enough to the identity. However, our preliminary numerical experiments on 3D non-rigid registration suggest that this limitation affects only very large deformations, and may not be problematic for image registration results. This novel setting is the infinite-dimensional analogous of the Log-Euclidean framework proposed in [2] for tensors. In this framework, usual Euclidean statistics can be performed on diffeomorphisms via their logarithms, whith excellent mathematical properties like inversion-invariance.

The rest of this paper is organized as follows. In the next Section, we present the Log-Euclidean framework for diffeomorphisms, which is closely linked to the notion of *one-parameter subgroups*. Then, we present two efficient algorithms to compute the exponential of a vector field and the logarithm of a diffeomorphism, which are exemplified on synthetic data. Finally, we apply our framework to non-linear registration results to compute a Log-Euclidean mean deformation between a 3D atlas and a dataset of 9 T_1 MR images of human brains.

2 A Log-Euclidean Framework for Diffeomorphisms

To generalize the notion of *principal logarithm* to diffeomorphisms, we will rely on its close link with *one-parameter subgroups*. Remarkably, one-parameter subgroups of diffeomorphisms take a simple form, and by analogy with the finite-dimensional case, we will define logarithms of diffeomorphisms as being the *infinitesimal generators* of these subgroups.

Exponential and One-Parameter Subgroups. Let $(G, .)$ be a group, which can typically be a matrix multiplicative group. Then a family of elements $(g(s))_{s\in\mathbb{R}}$ of G is a one-parameter subgroup of G if and only if $g(0)$ is the neutral element e of G and for all s, t: $g(s).g(t) = g(s+t)$. When such a subgroup belongs to a (finite-dimensional) Lie group and is continuous, it is also *differentiable*, and its derivative at 0, $\frac{dg}{ds}(0)$, is called its *infinitesimal generator*.

Examples of such subgroups are given by the exponential in the following way : $(\exp(s.M))_{s\in\mathbb{R}}$ is a one-parameter subgroup, where M belongs to the Lie algebra of G (i.e., is an element of its tangent space at e). Conversely, we have the remarkable result that *all* continuous one-parameter subgroups are precisely of the form $(\exp(s.M))_{s\in\mathbb{R}}$. For more details, see [12], Chap. V. As

a consequence, once one-parameter subgroups are identified in a Lie group, its exponential mapping is immediately known. This also yields the form taken by the logarithm, which is the (local) inverse of the exponential.

One-Parameter Subgroups of Diffeomorphisms. In the case of diffeomorphisms, what do continuous one-parameter subgroups look like? Quite intuitively, they are all obtained via the integration of *stationary* ODEs (also called *autonomous*), i.e ODE whose speed vector does not depend on time [13].

Let $\dot{x} = V(x)$ be a stationary ODE, where the vector field $V(x)$ is smooth enough. The *flow* associated to this ODE is the family of mappings $\Phi_V(.,s) : \mathbb{R}^n \to \mathbb{R}^n$ parameterized by a time parameter s, such that for a fixed x_0, $s \mapsto \Phi_V(x_0,s)$ is the unique solution of $\dot{x} = V(x)$ with initial condition x_0 at time 0.

Intuitively, for a fixed s, the flow $x \mapsto \Phi_V(x,s)$ gives the way the ambient space is deformed by the integration of the ODE during s units of time. Remarkably, the flow is a *one-parameter subgroup of diffeomorphisms*. This implies in particular that the deformations of space given at time 1 by $\Phi_V(.,1)$ are twice that observed at time 0.5 via $\Phi_V(.,0.5)$. $V(x)$ is the *infinitesimal generator* of this subgroup. Conversely, all continuous one-parameter subgroups of diffeomorphisms are flows of some stationary ODE. See [13] for more details.

Exponentials of Vector Fields and Logarithms of Diffeomorphisms. Here, we propose to define the *exponential* $\exp(V)$ of a (smooth) vector field $V(x)$ as the flow at time 1 of the stationary ODE $\dot{x} = V(x)$. This is the only possible definition generalizing to vector fields the equivalence between one-parameter subgroups and exponential that exists in the finite-dimensional case. In this setting, the *logarithm* $\log(\Phi)$ of a diffeomorphism Φ close enough to the identity is the unique vector field in a neighborhood of zero such that $\exp(V) = \Phi$.

In the finite-dimensional case, the existence and uniqueness of such a logarithm can be proved by showing that the exponential is continuously differentiable and that its differential mapping at zero is the identity, and is thus locally invertible. Here, in the infinite-dimensional case, let V be a regular vector field, let t be a non-zero scalar, and let $\Phi_{t.V}(x,1)$ be the flow at time 1 associated to $t.V(x)$. Then, a simple change of variable ($s \mapsto t.s$) shows that:

$$\lim_{t \mapsto 0} \frac{\Phi_{t.V}(x,1) - x}{t} = \lim_{t \mapsto 0} \frac{\exp(t.V)(x) - \exp(0)(x)}{t} = V(x),$$

which means intuitively that $\partial_V \exp(0) = V$, which indicates that the differential of the exponential is the identity, and that we have existence and uniqueness of the logarithm locally around the identity. To be entirely rigorous and to complete the proof, it would be necessary to define precisely the smoothness required of V, and to show the differentiability of the exponential in this infinite-dimensional setting. We are currently working on these quite technical details.

Log-Euclidean Statistics on Diffeomorphisms. On diffeomorphisms whose logarithm is well-defined, one can perform Euclidean operations, since these logarithms are simple vector fields. In particular, one can define a distance between these diffeomorphisms via a norm $\|.\|$ on vector fields: $\text{dist}(\Phi_1, \Phi_2) =$

$\|\log(\Phi_1) - \log(\Phi_2)\|$. Remarkably, this type of distance is *inversion-invariant*, since $\log(\Phi^{-1}) = -\log(\Phi)$. In the case of Hilbert norms, the point that minimizing the weighted sum of squared distances to the data is the *Log-Euclidean mean*, given by $\exp(\sum_i w_i \log(\Phi_i))$. This mean is inversion-invariant, and is also invariant with respect to the taking of square roots, since $\log(\Phi^{\frac{1}{2}}) = \frac{1}{2}\log(\Phi)$. More generally, in this setting, one can perform any kind of statistics on *diffeomorphisms* via *vectorial* statistics on their logarithms, which allows a straightforward generalization of classical analysis tools like Principal Component Analysis (PCA) on diffeomorphisms.

3 Computation of Exponentials and Logarithms

Fast Computation of Exponentials. Here, we present an efficient algorithm to compute exponentials of speed vectors on regular grids, which generalizes the Fast Polyaffine Transform we proposed in [1]. It is based on a non-linear generalization of the popular 'Scaling and Squaring' method, widely used to compute *matrix exponentials*. The basic idea is that the matrix exponential is much simpler to compute for matrices *close to zero*, for example using Padé approximants. In particular, one can compute very accurately $\exp\left(\frac{M}{2^N}\right)$ and obtain $\exp(M) = \exp(\frac{M}{2^N})^{2^N}$ by squaring recursively N times the result [7].

In the non-linear case, one can follow exactly the same steps as for matrices and generalize the 'Scaling and Squaring' to vector fields in the following way:

1. **Scaling step:** Divide $V(x)$ by a factor 2^N, so that $V(x)/2^N$ is close enough to zero (according to the level of accuracy desired).
2. **Exponentiation step:** $\Phi_V\left(\frac{1}{2^N}, .\right)$ is computed with a first-order explicit numerical scheme.
3. **Squaring step:** N recursive squarings of $\Phi_V\left(\frac{1}{2^N}, .\right)$ yield an accurate estimation of $\Phi_V(1, .)$ (only N *compositions of mappings* are used).

Intuitively, this means that we obtain the deformations at time 1 as a result of 2^N times the composition of the (very) small deformations observed at time $\frac{1}{2^N}$. This allows a fast integration of stationary ODEs on regular grids, compared to classical integrations based on fixed time-steps. To perform the composition of (sampled) deformations on regular grids, we use bi- or tri-linear interpolation, which guarantees the continuity of the interpolation in an simple way.

Fast Computation of Logarithms. Exactly as for exponentials, we use here a non-linear generalization of the popular 'Inverse Scaling and Squaring' method (ISS), widely used to compute *matrix logarithms*. Anew, the idea is that logarithms are much simpler to compute for matrices *close to the identity*, for instance with Padé approximants. To transform a matrix M so that it is closer to the identity, the ISS algorithm performs recursive computations of *square roots*. Then the identity $\log(M) = 2^N.\log(M^{2^{-N}})$ is used to compute $\log(M)$.

Let Φ be a diffeomorphism. To compute its logarithm V, we use the following non-linear generalization of the ISS algorithm:

1. **Scaling step:** Chose a scaling factor 2^N (according to the level of accuracy desired).
2. **Square rooting step:** $\Phi^{2^{-N}}$ is computed by N successive recursive takings of square roots.
3. **Computation of logarithm step:** $\log(\Phi)$ is given by $2^N.\log(\Phi^{2^{-N}})$, where $\log(\Phi^{2^{-N}})$ is simply estimated by $\Phi^{2^{-N}} - Id$ (where Id is the identity).

In order to compute square roots, we perform here a gradient descent on the functional $E_{\mathrm{SQRT}}(T) = \frac{1}{2}.\int_\Omega \|T \circ T - \Phi\|^2(x)dx$, with $\frac{1}{2}.(\Phi - Id)$ as initialization. The (L^2) gradient of this energy is the following:

$$\nabla E_{\mathrm{SQRT}}(T) = (DT^t) \circ T.(T \circ T - \Phi) + |\det(D(T^{-1}))|(T - \Phi \circ T^{-1}),$$

where DT is the Jacobian of T and where '.' and 'M^t' are matrix multiplication and transposition. Exactly as in the matrix case, this requires the inversion of T, which we also perform by gradient descent on $E_{\mathrm{INV}}(T) = \frac{1}{2}.\int_\Omega \|\Phi \circ T - Id\|^2(x)dx$, with an initialization of $-(\Phi - Id)$. The (L^2) gradient is given by:

$$\nabla E_{\mathrm{INV}}(T) = (D\Phi^t) \circ T.(\Phi \circ T - Id).$$

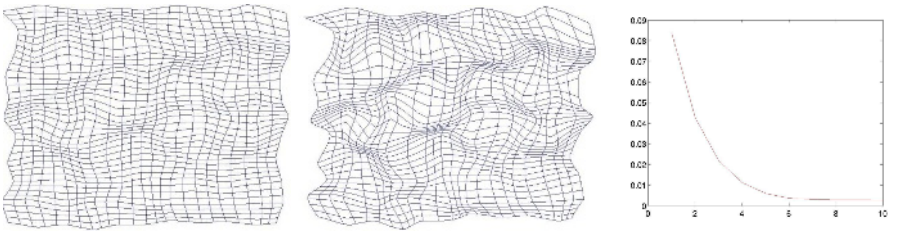


Fig. 1. Fast computation of the exponential of a random vector field. From left to right: the two last iterations (scaling of 2^8) of our fast computation of the exponential and the evolution of the average relative accuracy with N. Note how deformations increase exponentially from one iteration to another. The relative accuracy obtained on average converges toward of 0.3%, which is approximately obtained for $N = 7$.

Experiments on 2D Synthetic Data. In these experiments, we evaluate the accuracy of our algorithms on 2D random deformations, sampled on a 40×40 regular grid. To generate a random vector field, random displacements are computed on a 11×11 regular grid (Gaussian white noise of standard deviation 0.2, coordinates of grid points in $[-5, 5]$) , which are smoothly interpolated on the finer grid using bilinear interpolation.

We measure the accuracy of the fast computation of the exponential by computing the relative difference with respect to a very accurate estimation of the continuous transformation, obtained by a classical integration (i.e., with fixed time step, here 2^{-8}) of the stationary ODE associated to the random vector

field, for each voxel of the grid. Fig. 1 presents the results, which show that a typical accuracy on average of 0.3% percent can be obtained with 7 squarings.

To evaluate the accuracy of the fast computation of the logarithm, we first estimate the exponential of the random vector field with our fast algorithm (with 8 squarings). Then, the logarithm of this diffeomorphism is computed and is compared to the original vector field (10 iterations are used for each gradient descent). Fig. 2 presents the results, which show that an accuracy of 3% percent can be obtained on average with 6 recursive computations of square roots.

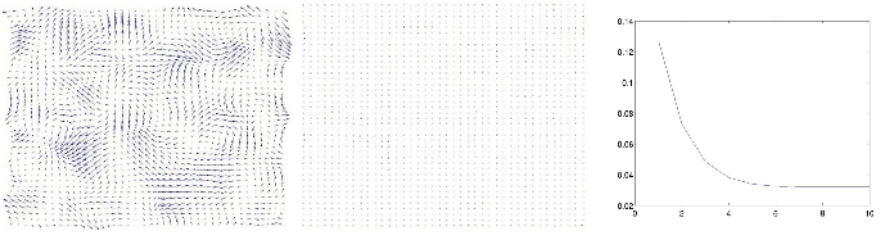


Fig. 2. Fast computation of the logarithm of a random deformation. From left to right: random vector field (logarithm), then difference between original and estimated logarithms (for $N = 8$), evolution of accuracy with N. The relative accuracy obtained on average converges toward of 3%.

4 Statistics on 3D Diffeomorphisms

In this Section, we compute Log-Euclidean statistics on the registration results obtained between a $256 \times 256 \times 60$ artificial T_1 MR image of a human brain (coming from the BrainWeb [1]), referred to here as the 'atlas', and a dataset of 9 T_1 images. The images of the dataset are first globally aligned with the atlas using a robust block-matching affine registration algorithm. Then, a fine registration is performed using a dense transformation registration algorithm guaranteeing the invertibility of the global deformations [11]. Statistics are performed on the deformations from the atlas to each of the subjects' geometry obtained in the non-linear registration step.

To compute logarithms of diffeomorphisms and exponential of vector fields, we used the fast algorithms presented in the former section. A scaling of 2^8 is used in both cases, and a maximal number of 10 iterations is used during each gradient descent.On the $256 \times 256 \times 60$ grid, using a Intel M processor 2.13GHz with 1 Go of RAM, an exponential was computed in typically $30s$ and a logarithm in 60 minutes.

Fig. 3 presents the results obtained on mean global deformations. We see in particular that the Log-Euclidean mean deformation tends to increase the size of ventricles, which are smaller in the atlas than in the dataset. We also compared the Log-Euclidean mean to the Euclidean mean of displacements, which is not

[1] Web site: http://www.bic.mni.mcgill.ca/brainweb/

guaranteed to be invertible. In this example, both means are quite close to each other, although locally, one can observe in the region of large mean deformations relative differences of the order of 30%, for example in the ventricles.

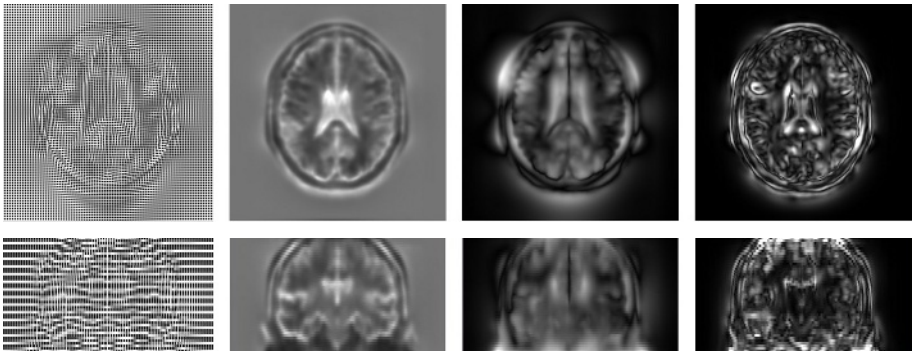


Fig. 3. Deformation of atlas with Log-Euclidean mean diffeomorphism. From left to right, images of: regular grid deformed by Log-Euclidean mean deformations, Jacobian of Log-Euclidean mean deformations, norm of Log-Euclidean mean deformations, and finally norm of difference between mean Euclidean and Log-Euclidean deformations of atlas. The largest mean deformations are observed in the ventricles, on the cortex and the skull; this is due to the anatomical differences between the atlas and the population. On the Jacobian map, we see in particular that high values are obtained in the ventricles: this comes the fact that the atlas has on average smaller ventricles than in the population. In this example, both means are quite close to each other, although locally, one can observe in the region of large mean deformations relative differences of the order of 30%, for example in the ventricles.

5 Conclusions and Perspectives

In this article, we have presented an infinite-dimensional generalization to diffeomorphisms of our Log-Euclidean framework for tensors [2], based on the generalization of the notion of *principal logarithm* to invertible deformations. This logarithm is a simple 3D vector field, and is well-defined for diffeomorphisms close enough to the identity. In the set of diffeomorphisms whose logarithm is well-defined, one can perform Euclidean operations, since these logarithms are simple vector fields. As a consequence, this framework can be used to perform *vectorial* statistics on diffeomorphisms which are always preserve the invertibility constraint (contrary to Euclidean statistics on displacement fields), and are also invariant with respect to inversion.

Furthermore, we have also presented two efficient algorithms which successfully generalize to the non-linear case two popular algorithms used to compute the matrix exponential and logarithm. In practice, this allows for example the computation of Log-Euclidean means of 3D geometrical transformations, as we have shown in the context of a registration experiment between an atlas an a

database of 9 subjects. This type of statistics could prove very useful to quantitatively compare registration algorithms, or to constrain them. Also, in the domain of image and shape statistics, the Log-Euclidean framework for diffeomorphisms could provide an interesting setting to build models with a constant topology.

References

1. V. Arsigny, O. Commowick, X. Pennec, and N. Ayache. A Log-Euclidean polyaffine framework for locally rigid or affine registration. In *Proc. of WBIR'06*, volume 4057 of *LNCS*, pages 120–127, Utrecht, The Netherlands, July 2006.
2. V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Fast and simple calculus on tensors in the Log-Euclidean framework. In *MICCAI (1)*, pages 115–122, 2005.
3. M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. Jour. Comp. Vis.*, 61(2):139–157, 2005.
4. V. Camion and L. Younes. Geodesic interpolating splines. In *Proc. of EMMCVPR'01*, LNCS 2134, pages 513–527, 2001.
5. C. Chefd'hotel, G. Hermosillo, and O. Faugeras. Flows of diffeomorphisms for multimodal image registration. In *Proc. of ISBI*, 2002.
6. O. Commowick, R. Stefanescu, P. Fillard, V. Arsigny, N. Ayache, X. Pennec, and G. Malandain. Incorporating statistical measures of anatomical variability in atlas-to-subject registration for conformal brain radiotherapy. In *Proc. of MICCAI'2005 (II)*, LNCS, pages 927–934, 2005.
7. N. J. Higham. The scaling and squaring method for the matrix exponential revisited. *SIAM J. Matrix Anal. Appl.*, 26(4):1179–1193, 2005.
8. S. Marsland and C. J. Twining. Constructing diffeomorphic representations for the groupwise analysis of nonrigid registrations of medical images. *IEEE Trans. Med. Imaging*, 23(8):1006–1020, 2004.
9. X. Pennec, R. Stefanescu, V. Arsigny, P. Fillard, and N. Ayache. Riemannian elasticity: A statistical regularization framework for non-linear registration. In *Proc. of MICCAI 2005 (II)*, LNCS, pages 943–950, 2005.
10. D. Rueckert, A. F. Frangi, and J. A. Schnabel. Automatic construction of 3D statistical deformation models of the brain using non-rigid registration. *IEEE TMI*, 22(8):1014–1025, 2003.
11. R. Stefanescu, X. Pennec, and N. Ayache. Grid powered nonlinear image registration with locally adaptive regularization. *MedIA*, 8(3):325–342, September 2004.
12. S. Sternberg. *Lectures on Differential Geometry*. Prentice Hall Mathematics Series. Prentice Hall Inc., 1964.
13. M. Tenenbaum and H. Pollard. *Ordinary Differential Equations*. Dover, 1985.
14. A. Trouvé. Diffeomorphisms groups and pattern matching in image analysis. *International Journal of Computer Vision*, 28(3):213–221, 1998.
15. M. Vaillant, M. Miller, L. Younes, and A. Trouvé. Statistics on diffeomorphisms via tangent space representations. *NeuroImage*, 23:S161–S169, 2004.

Nonrigid 3D Brain Registration Using Intensity/Feature Information

Christine DeLorenzo[1], Xenophon Papademetris[1,2], Kun Wu[3], Kenneth P. Vives[3], Dennis Spencer[3], and James S. Duncan[1,2]

[1] Departments of Electrical Engineering,
[2] Diagnostic Radiology, and
[3] Neurosurgery, Yale University, P.O. Box 208042
New Haven CT 06520-8042, USA
{christine.delorenzo, xenophon.papademetris, kun.wu, kenneth.vives, dennis.spencer, james.duncan}@yale.edu

Abstract. The brain deforms non-rigidly during neurosurgery, preventing preoperatively acquired images from accurately depicting the intraoperative brain. If the deformed brain surface can be detected, biomechanical models can be applied to calculate the resulting volumetric deformation. The reliability of this volumetric calculation is dependent on the accuracy of the surface detection. This work presents a surface tracking algorithm which relies on Bayesian analysis to track cortical surface movement. The inputs to the model are 3D preoperative brain images and intraoperative stereo camera images. The addition of a camera calibration optimization term creates a more robust model, capable of tracking the cortical surface in the presence of camera calibration error.

1 Introduction

The use of Surgical Navigation Systems (SNS) has greatly aided neurosurgical procedures. These systems provide real-time visualization of surgical tool positions relative to preoperative patient brain images, allowing surgeons to locate pathologic structures and other surgical targets. As SNS use becomes more prevalent in varied surgical procedures, the accuracy and reliability of these systems must be demonstrated. Soft tissue movement due to gravity, loss of blood and cerebrospinal fluid (CSF), and the action of certain medications can lead to misalignment between preoperative brain scans and the intraoperative brain [4, 9]. Therefore, accounting for this brain shift is necessary to realize the effectiveness of these navigation systems and to allow surgeons to confidently excise tissue, especially near functionally eloquent areas.

Volumetric images revealing the location and amount of brain shift can be taken intraoperatively. However, surgical constraints limit both the time per acquisition as well as prevent multi-modality imaging and use of certain modalities such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET). The effect of these restrictions is decreased spatial resolution, lower contrast and the loss of meaningful information [14, 15]. Additionally,

R. Larsen, M. Nielsen, and J. Sporring (Eds.): MICCAI 2006, LNCS 4190, pp. 932–939, 2006.

intraoperative magnetic resonance imaging (iMRI) and computed tomography (iCT) are prohibitively expensive [15] for most hospitals.

Since the multitude of factors affecting brain shift, coupled with patient variability, make it impossible to predict brain shift a priori, it becomes necessary to characterize brain movement using sparse intraoperative data to update the rich preoperative information. Warfield et al. [3, 14] warped preoperative images containing projected atlas information through the use of surface data acquired by iMRI and a volumetric biomechanical model. Their study focused on the visualization of the corticospinal tract, which is not observable with conventional MRI, and indicated visually appealing results can be obtained in real time. Several groups use commercially available laser range scanners (LRS) to detect the cortical surface. Registration can then be performed between the acquired depth maps. Audette et al. [1] used a non-rigid iterative closest point (ICP) algorithm for this registration in which the initial alignment must be done manually. This work did not take advantage of any surface features such as blood vessels or sulci. In [10], the LRS acquired both a range and texture map, allowing the non-rigid registration to be performed in the 2D texture space, which uniquely corresponds to a depth map.

Nakajima et al. [7] used cortical surface vessels for registration of 2D intraoperative images with 3D postcontrast MRI images. Cortical vessels were chosen as landmarks due to their similarity with sulcal and gyral patterns and ease of identification in phase contrast MRI. Sun et. al [12] also used cortical surface vessels to guide surface registration in a phantom study. Vessels were simulated by metal wires and imaged using CT. Intraoperative surface information was acquired using a single charge-coupled device (ccd) camera mounted on an operating microscope. To perform stereo reconstructions of the surface, two images were acquired at different microscope positions, which were tracked optically. Sun et al. [13] later replaced this with a stereo camera mounted microscope. Škrinjar [11] et al. used a similar stereo camera setup. His work combined the model deformation and stereo reconstruction. However, neither study used feature matching with the stereo reconstruction.

This work extends that of previous authors by performing both intensity and feature based nonrigid registration of cortical brain surface using a stereo camera setup. Since the reliability of a system based on cameras is highly dependent on the accuracy of the camera calibration, this work seeks to correct for any calibration inaccuracies and update the calibration matrices with the surface deformation calculation. This framework is both convenient and versatile. (Digital cameras are either present or can easily be transported to the operating room.) By using sulcal grooves and not cortical vessels as features, this algorithm does not require patients to ingest contrast or endure more than the standard preoperative imaging. And, it takes full advantage of the feature information to improve the deformation prediction while also correcting for any calibration inaccuracies, which would limit the reliability of the result. In this way, many of the advantages of the above-proposed models are captured in one algorithm.

2 Method

A surface detection algorithm which calculates the most likely intraoperative brain surface deformation using Bayesian analysis was outlined in [2]. This work developed a model predicting the dense displacement field of the cortical surface, given intraoperative images and surface feature locations. This displacement field, $\underline{\hat{U}_{dense}}$, was found to be:

$$\underline{\hat{U}_{dense}} = \arg\max_{\underline{U_{dense}}} p(\underline{U_{dense}}|\mathbf{I}, \mathbf{K}, \underline{C}, S^U, \mathbf{A}) \tag{1}$$

where $\underline{U_{dense}}$ is a possible dense displacement field. $\mathbf{A} = [A_0, A_1]$, where A_0, A_1 are the camera calibration matrices acquired by the left (0) and right (1) cameras. $\mathbf{I} = [I_0, I_1]$, where I_0, I_1 are the intraoperative images acquired by the left (0) and right (1) cameras. $\mathbf{K} = [K_0, K_1]$, where K_0, K_1 are matrices in which each column is the vector representation of a 2D sulcal groove traced by an expert user from images I_0 and I_1, respectively. $\underline{C}$ is a matrix in which each column is the vector representation of a 3D preoperative sulcus position, and S^U is the original (preoperative) cortical surface, extracted from the preoperative MRI.

This method yielded surface deformation predictions with submillimeter mean accuracy both in simulation and phantom experiments. The camera calibration matrices were critical for taking advantage of both the surface intensities, as acquired from the stereo cameras, and the feature alignment, which requires a projection from 3D space to the image plane. However, ideal calibrations are not feasible in many real world situations [6], especially in a busy operating room where time is of the essence. Therefore, we propose updating the camera calibration calculation in the region of interest using the extra scene information obtained from cortical feature information obtained during neurosurgery.

The updated model can be written as follows:

$$(\underline{\hat{U}_{dense}}, \hat{\mathbf{A}}) = \arg\max_{\underline{U_{dense}}, \mathbf{A}} p(\underline{U_{dense}}, \mathbf{A}|\mathbf{I}, \mathbf{K}, \underline{C}, S^U) \tag{2}$$

Equation (2) can be solved using an iterative scheme that updates both the camera calibration and the dense displacement field calculation at every iteration.

$$\underline{\hat{U}^k_{dense}} = \arg\max_{\underline{U_{dense}}} p(\underline{U_{dense}}|\mathbf{I}, \mathbf{K}, \underline{C}, S^U, \mathbf{A}^{k-1}) \tag{3}$$

$$\hat{\mathbf{A}}^k = \arg\max_{\mathbf{A}} p(\mathbf{A}|\mathbf{I}, \mathbf{K}, \underline{C}, S^U, \underline{U_{dense}}^{k-1}) \tag{4}$$

where k is the iteration number.

Maximizing the expressions in equations (3) and (4) is equivalent to maximizing the log of the posterior probabilities. The expansion of equation (3) can be found in [2].

2.1 Camera Calibration Optimization

Looking at equation (4), the most likely camera calibration, $\underline{\hat{\mathbf{A}}}$, will best match 3D cortical surface points to their corresponding 2D image locations. The imaged sulci provide accurate feature information from which correspondence can be reliably determined. Since the brain deformation is small relative to the brain volume, the sulcal positions should not vary greatly between frames. Because of this, the projection of sulci from the 3D preoperative brain surface, $\underline{C}$, should be in close proximity to the intraoperatively imaged sulci, $\underline{\mathbf{K}}$. Using this information and further simplification, the determination of $\underline{\hat{\mathbf{A}}}^{k}$ can be reduced to the maximum argument over $\underline{\mathbf{A}}$ of:

$$\log\left[p(\underline{\mathbf{A}}|\underline{\mathbf{K}}, \underline{C}, \underline{U_{dense}}^{k-1})\right] \tag{5}$$

Assuming independence of the stereo cameras, equation (5) can be written as:

$$\log\left[p(A_0|\underline{K_0}, \underline{C}, \underline{U_{dense}}^{k-1})\right] + \log\left[p(A_1|\underline{K_1}, \underline{C}, \underline{U_{dense}}^{k-1})\right] \tag{6}$$

To solve equation (6), it is assumed that the camera calibration matrices from the the previous iteration are known. For iteration k = 1, the k-1 calibration matrices are the initial matrices obtained from calibration procedure described in [2]. According to [5], assuming the measurement error in calculating the calibration matrices is Gaussian, the reconstruction error will be Gaussian. Therefore, equation (6) reduces to:

$$\log\left[p(\underline{\mathbf{A}}|\underline{\mathbf{K}}, \underline{C}, \underline{U_{dense}}^{k-1})\right] = -\eta_{0,1} \int d[\underline{K_{0,1}} - (A_{0,1} \cdot (\underline{C} + \underline{U^{C}_{dense}})]dS \tag{7}$$

where d is a distance metric, $\underline{U^{C}_{dense}}$ represents the deformation field restricted to the sulci, and $\eta_{0,1}$ are constants.

Whereas the search for the displacement field can take advantage of the image intensities, it would be difficult to incorporate these intensities into a search for the optimum camera calibration matrix. This is why the use of features is critical to this process.

2.2 Data Acquisition

Noncontrast preoperative MRI images were obtained on a 1.5 T MR scanner. The preoperative cortical surface was extracted and 3D sulci positions were located manually. The camera calibration was performed using the method outlined in [2]. Images (1024 x 768) were acquired from a mounted ccd camera in the OR, which could be repositioned by the surgeon. For validation purposes, the 3D locations of various points on the exposed cortical surface were selected with a BrainLAB pointer tool (BrainLAB AG, Heimstetten, Germany), and saved using the Vector Vision Cranial software [2, 8].

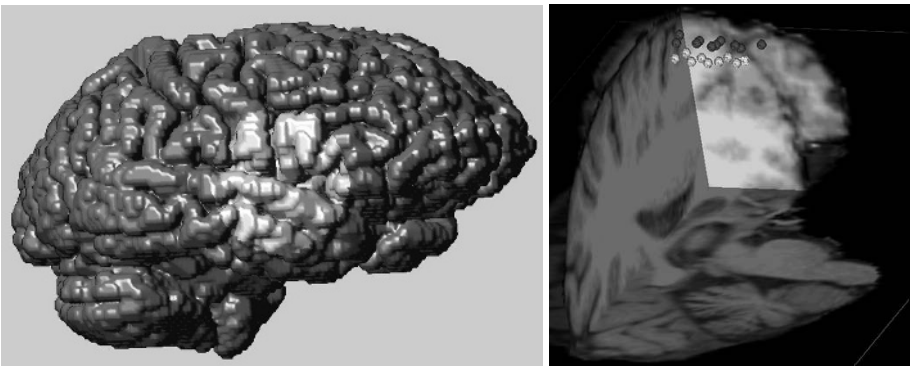


Fig. 1. The left image depicts the extracted 3D brain surface. The craniotomy was performed in the highlighted section. The right image shows initial (dark spheres) and final (light spheres) brain surface positions as acquired in the OR for algorithm validation. The hardware and software required for point acquisition are explained in [8].

3 Results

Seven patients undergoing stage I epilepsy surgery, in which intracranial electrodes are inserted to monitor the patient during and between seizures, were chosen for brain shift tracking. Though the craniotomies varied in position, the size of the skull opening, as well as the surgical time, were similar across all patients. One sample case is explored below.

Figure 1 shows the extracted preoperative brain surface and the 3D point locations selected during surgery for validation. The 3D manually-selected preoperative sulcal locations, smoothed extracted surface and intraoperative pictures with outlined sulci (Figure 2) were inputs to the surface detection algorithm. The algorithm took 23.5 minutes to run on a 3 GHz Pentium 4 computer. As seen qualitatively in Figure 3, the algorithm closely predicts the deformed cortical surface position. The coloring on the 3D patches represents greylevel intensities projected on the surface from one of the intraoperative images. Due to the misalignment caused by the soft tissue deformation, most of the initial surface patch is not in the camera view. However, the actual and predicted surface show the correct intensities, reflecting their accurate location relative to the OR camera.

The mean reconstruction error was 1.01 mm when the calibration optimization was used. As can be seen in Table 1, without the camera calibration correction, the surface matching result is much worse. This makes intuitive sense because calibration accuracy determines the effectiveness of both the image intensities and the projected feature information. In this case, the optimized calibration matrices were less than 2% different from the initial matrices. This small difference, which could have been the result of fitting error or slight camera movement, created a significant difference in the algorithm result.

Table 1. Surface Deformation Tracking Results. The actual surface displacement was measured by the BrainLAB VV Cranial software and compared to the displacement predicted by the algorithm, with and without camera calibration optimization. When the camera calibration matrices are inaccurate, the intensity and feature information cannot be reliably used to track the surface.

	mean (mm)	*max (mm)*	*std (mm)*
surface displacement	7.49	8.38	0.61
algorithm error, no calibration correction	6.80	11.01	1.98
algorithm error, calibration correction	1.01	1.97	0.60

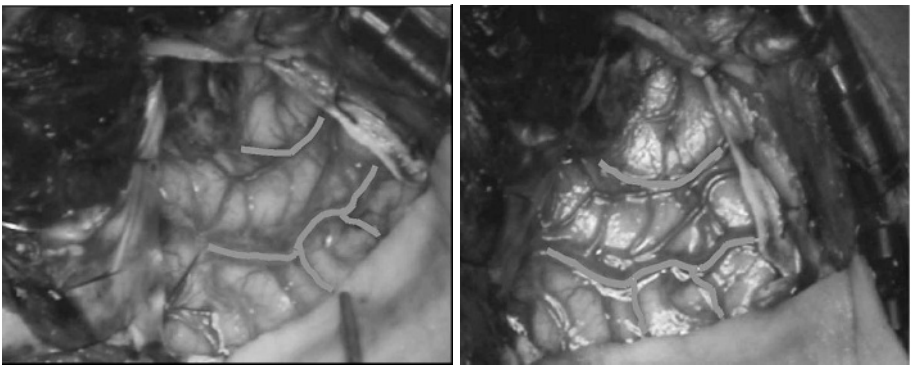


Fig. 2. Stereo camera intraoperative images acquired approximately three hours after surgery began. The sulci used as the feature information are outlined.

4 Discussion

During neurosurgery, it is essential to accurately locate pathologic structures within the brain. Due to the similarity in appearance between these pathologic regions and healthy tissue, neurosurgeons often rely on SNS, though these systems cannot currently compensate for intraoperative brain shift. Biomechanical models can be used to predict this nonrigid brain shift, if driven by accurate intraoperative information, such as cortical surface deformation. This work made use of a stereo camera setup to track this cortical surface displacement. Future model refinement will focus on application of the model to a greater number of surgical cases, updating model parameters for optimum accuracy, as well as using the acquired surface results with biomechanical models to track the volume deformation.

Though it has been previously shown that intraoperative images can be used to drive surface tracking models, the results have always been susceptible to camera calibration error. This is especially true in patient studies in which OR time, and therefore camera calibration time, is limited. As seen in the above test case, the surface tracking algorithm would have failed without the calibration

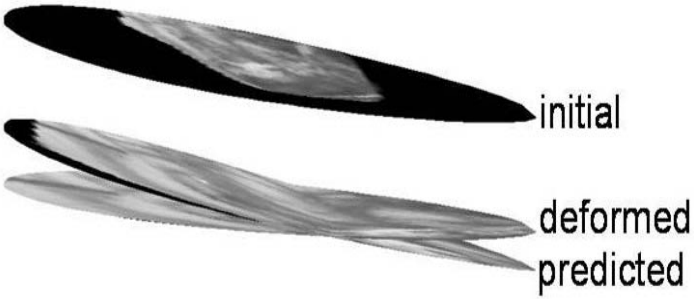


Fig. 3. The color of the surfaces patches was found by backprojecting the image intensities to the surface using the camera calibration matrices. The cortical surface sinks during surgery from its initial position, out of the intraoperative camera view, to the deformed position. The surface tracking algorithm takes advantage of both the image intensities and the feature information to closely determine the final cortical position.

optimization, though the calibration matrices were within two percent of their correct values. The addition of the camera calibration optimization term, however, yielded significantly better results. The combined algorithm predicted the surface displacement with an error less than, or within the range of, previously published surface deformation studies [3, 10, 13]. This work shows that by taking advantage of the cortical feature information, the surface displacement can be tracked in the presence of calibration inaccuracies.

References

1. M. A. Audette, K. Siddiqi, F. P. Ferrie, and T. Peters: An integrated range-sensing, segmentation and registration framework for the characterization of intra-surgical brain deformations in image-guided surgery. Computer Vision and Image Understanding, vol. **89** (2003) 226-251
2. C. DeLorenzo, X. Papademetris, K. P. Vives, D. Spencer and J. S. Duncan: Combined feature/intensity-based brain shift compensation using stereo guidance. *In International Symposium on Biomedical Imaging*, (ISBI 2006) 335–338
3. M. Ferrant, A. Nabavi, B. Macq, P. M. Black, F. A. Jolesz: Serial registration of intraoperative MR images of the brain. Medical Image Analysis, vol. **6** No. **4** (2002) 337-360
4. D. T. Gering, A. Nabavi, R. Kikinis, W. E. L. Grimson, N. Hata, P. Everett, F. Jolesz, and W. M. Wells: An integrated visualization system for surgical planning and guidance using image fusion and interventional imaging. *In Medical Image Computing and Computer-Assisted Intervention*, (MICCAI 1999) 809–819
5. S. K. Kopparapu and P. Corke: The effect of measurement noise on intrinsic camera calibration parameters. Proceedings of the 1999 IEEE International Conference on Robot and Automation (May 1999) 1281–1286
6. M. Machacek, M. Sauter, and T. Rösgen: Two-step calibration of a stereo camera system for measurements in large volumes. Measurements in Science and Technology, vol. **14** (2003) 1631–1639

7. S. Nakajima, H. Atsumi, R. Kikinis, T. Moriarty, D. Metcalf, F. Jolesz, and P. Black: Use of cortical surface vessel registration for image-guided neurosurgery. Neurosurgery, vol. **40** (1997) 1201-1210
8. X. Papademetris, K. P. Vives, M. DiStasio, L. H. Staib, M. Neff, S. Flossman, N. Frielinghaus, H. Zaveri, E. J. Novotny, H. Blumenfeld, R. T. Constable, H. Hetherington, R. Duckrow, S. S. Spencer, D. Spencer and J. S. Duncan: Development of a research interface for image guided intervention: Initial application to epilepsy Neurosurgery. *In International Symposium on Biomedical Imaging*, (ISBI 2006) 490–493
9. D. W. Roberts, A. Hartov, F. E. Kennedy, M. I. Miga, and K. D. Paulsen: Intraoperative brain shift and deformation: A quantitative analysis of cortical displacement in 28 cases. Neurosurgery, vol. **43** No. **4** (October 1998) 749–758
10. T. K. Sinha, V. Duay, B. M. Dawant, and M. Miga: Cortical shift tracking using a laser range scanner and deformable registration methods. *In Medical Image Computing and Computer-Assisted Intervention*, (MICCAI 2003) 166–174
11. O. Škrinjar, and J. Duncan: Stereo-guided volumetric deformation recovery. *In International Symposium on Biomedical Imaging*, (ISBI 2002) 863–866
12. H. Sun, D. W. Roberts, A. Hartov, K. Rick, and K. D. Paulsen: Using cortical vessels for patient registration during image-guided neurosurgery - A phantom study.. Proceedings of SPIE, vol. **5029** (May 2003) 183–191
13. H. Sun, H. Farid, K. Rick, A. Hartov, D.W. Roberts, and K.D. Paulsen: Estimating cortical surface motion using stereopsis for brain deformation models. *In Medical Image Computing and Computer-Assisted Intervention*, (MICCAI 2003) 794–801
14. S. K. Warfield, F. Talos, A. Tei, A. Bharatha, A. Nabavi, M. Ferrant, P. M. Black, F. A. Jolesz, and R. Kikinis: Real-time registration of volumetric brain MRI by biomechanical simulation of deformation during image guided neurosurgery. Computing and Visualization in Science. (2002) 3–11
15. C. R. Wirtz, M. M. Bonsanto, M. Knauth, V. M. Tronnier, F. K. Albert, A. Staubert, and S. Kunze: Intraoperative magnetic resonance imaging to update interactive navigation in neurosurgery: Method and preliminary experience. Computer Aided Surgery, vol. **2** (1997) 172–179

Author Index