Cristian A. Linte   Elvis C.S. Chen
Marie-Odile Berger   John T. Moore
David R. Holmes III (Eds.)

Augmented Environments for Computer-Assisted Interventions

7th International Workshop, AE-CAI 2012
Held in Conjunction with MICCAI 2012
Nice, France, October 2013, Revised Selected Papers

Springer

Lecture Notes in Computer Science 7815

Cristian A. Linte   Elvis C.S. Chen
Marie-Odile Berger   John T. Moore
David R. Holmes III (Eds.)

Augmented Environments for Computer-Assisted Interventions

Springer

Volume Editors

Cristian A. Linte
David R. Holmes III
Mayo Clinic, Biomedical Imaging Resource
200 First Street SW
Rochester, MN 55905, USA
E-mail:{linte.cristian, holmes.david3}@mayo.edu

Elvis C.S. Chen
John T. Moore
Robarts Research Institute
Imaging Research Laboratories
100 Perth Dr.
London, ON N6A 5K8, Canada
E-mail: chene@imaging.robarts.ca
E-mail: jmoore@robarts.ca

Marie-Odile Berger
INRIA and University of Nancy
615 rue du Jardin Botanique
54603 Nancy, France
E-mail: marie-odile.berger@inria.fr

Preface

As organizers of the 7$^{\text{th}}$ MICCAI Workshop on Augmented Environments for Computer-Assisted Interventions (AE-CAI 2012), and editors of this volume, it is our pleasure to present to you the second edition of the Springer LNCS AE-CAI proceedings. This volume includes selected papers from those presented at the AE-CAI 2012 workshop held in conjunction with the Medical Image Computing and Computer-Assisted Interventions (MICCAI) 2012 conference, held on October 5, 2012, in beautiful Nice, Cote d'Azur, France.

AE-CAI 2012 was the seventh in a series of workshops formerly known as Augmented Medical Imaging and Augmented Realities for Computer-Aided Surgery (AMI-ARCS), following AMI-ARCS 2003, 2004, 2006, 2008, and 2009, AE-CAI 2011, and the Tutorial on Augmented Reality in 2007. Over the past several years, the satellite workshops and tutorials at MICCAI have experienced increased popularity. Their registration numbers in 2012 (over 1,000 registrants) competed with the registration numbers for the main conference ($\sim$ 1,100 registrants). The 2012 edition of AE-CAI reached just under 50 registrants, not including the members of the Organizing and Program Committees. There was also an impressive number of attendees ($\sim$ 60 attendees), making AE-CAI one of the best received and best attended workshops at MICCAI 2012.

The 2012 edition of the workshop was a joint effort between the Biomedical Imaging Resource at Mayo Clinic (Rochester, MN, USA), the Imaging Research Laboratories at Robarts Research Institute (London, ON, Canada), and the University of Nancy (Nancy, France). These groups have had a long-standing tradition in medical image analysis and image-guided intervention research and have been at the forefront of technological and translational development in this field for the past three decades. In addition, a Program Committee consisting of 40 international experts served as reviewers for the submitted papers.

In medical imaging, virtual and augmented reality environments aim to provide the physician with enhanced visualization and perception of the patient, either by fusing various imaging modalities or by presenting image-derived information overlaid on the physician's view, establishing a direct relation between the image and the patient. The objective of the AE-CAI workshop is to attract scientific contributions that offer solutions to the technical problems in the area of augmented and virtual environments for computer-assisted interventions, and to provide a venue for the dissemination of papers describing both complete systems and clinical applications. AE-CAI 2012 attracted researchers in computer science, biomedical engineering, physics, and clinical medicine engaged in the development of virtual and augmented environments for medical image visualization and image-guided interventions. As clinical impact is a core

element of the workshop, authors described their work in the context of necessary developments, research questions, and potential obstacles that must be overcome to enable and facilitate effective translation of the proposed work from lab to clinic. The scientific program focused on the dissemination of innovative research in image registration and fusion, calibration, visualization and 3D perception, hardware and optical design, real-time implementation, as well as validation, clinical applications, and clinical evaluation.

This workshop has traditionally included keynote addresses from leading figures in the field and 2012 was no exception. Richard Robb (Mayo Clinic, Rochester, MN, USA) spoke on the importance of building technology centered on best addressing the patient needs and gave several compelling examples of medical imaging and image-guided intervention breakthroughs that have served this purpose. René Anxionnat (University of Strasbourg, France) described state-of-the-art developments in augmented fluoroscopy and its applications in vascular neuroradiology for intraoperative guidance. The workshop concluded with an interactive panel discussion on the encountered successes and challenges of augmented, virtual, and image-guided intervention platforms toward their clinical translation and implementation.

AE-CAI 2012 attracted 22 paper submissions from 10 countries. The submissions were distributed for review to the Program Committee, ensuring all potential conflicts of interest were eliminated. Each paper was critically evaluated by at least four reviewers, who provided detailed critiques and constructive comments to the authors and workshop editorial board. Based on the reviews, 10 papers were selected for oral presentation (according to the reviewers' scores and suitability of the work for oral/poster presentation) and 8 were selected as poster presentations. The authors were asked to revise their initial submissions for publication in the Electronic Workshop Proceedings – made available to the workshop registrants on USB flash drives. Following the workshop, 14 of the papers presented at the meeting (representing 64% of all submitted papers) were selected for consideration of publication in this Springer LNCS volume, along with two editorials from our keynote speakers. The authors were asked to include both the reviewers' suggestions as well as the feedback received during the workshop, and resubmit their manuscripts, along with their response to reviewers, for a final review process toward publication in this volume. Each revised paper was reviewed by the volume editors to ensure that all reviewers' comments were properly addressed and incorporated into the final version of the manuscripts enclosed in this collection.

On behalf of the AE-CAI 2012 Organizing Committee, we would like to extend our sincere thanks to all Program Committee members for providing detailed and timely reviews of the submitted manuscripts. We greatly appreciate their support in providing valuable feedback. We also thank all authors, presenters, and attendees at AE-CAI 2012 for their scientific contribution, enthusiasm, and support. Lastly, we would like to express our most sincere thanks to

Northern Digital Inc. (NDI, Waterloo, ON, Canada) for their generous sponsorship, which enabled us to award the top oral and poster presentations at AE-CAI 2012. We hope that you all will enjoy reading this volume. We look forward to your continuing contribution and participation in the MIAR & AE-CAI joint events to be hosted in 2013 in Nagoya, Japan.

<div align="right">

Cristian A. Linte
Elvis C.S. Chen
Marie-Odile Berger
John T. Moore
David R. Holmes III

</div>

Organization

Organizing Committee

Cristian A. Linte	Mayo Clinic, USA
Elvis C.S. Chen	USA Robarts Research Institute, Canada
Marie-Odile Berger	University of Nancy, France
John T. Moore	Robarts Research Institute, Canada
David R. Holmes III	Mayo Clinic, USA

Associate Consultants

Eddie Edwards	Imperial College London, UK
Kensaku Mori	Nagoya University, Japan
Hongen Liao	University of Tokyo, Japan

Program Committee

Adrien Bartoli	Universite d'Auvergne
Michael Figl	Medical University of Vienna
Theo van Walsum	Erasmus Medical Center
Hongen Liao	University of Tokyo
Wolfgang Birkfellner	Medical University of Vienna
Johann Hummel	Medical University of Vienna
Tobias Blum	Technical University of Munich
Nassir Navab	Technical University of Munich
Jan Klein	Fraunhofer Mevis
Ron Kikinis	Harvard University
Stefan Wersag	Fraunhofer IGD
Bernhard Preim	University of Madgeburg
Rudy Lapeer	University of East Anglia
David Kwartowitz	Clemson University
Gabor Fichtinger	Queen's University
Bernhard Kainz	Graz University of Technology
Kensaku Mori	Nagoya University
Su-Lin Lee	Imperial College London
Marta Kersten	McGill University
Kevin Cleary	Children's National Medical Center
Ken Masamune	University of Tokyo
Stephane Nicolau	IRCAD Research Center
Henry Fuchs	University of North Carolina

Terry Peters   Robarts Research Institute
Kate Alicia Gavaghan   University of Bern
Jannick Rolland   University of Rochester
Coert Metz   Erasmus Medical Center
Gian-Luca Mariottini   University of Texas Arlington
Philip Pratt   Imperial College London
Amber Simpson   Vanderbilt University
Stamatia Giannarou   Imperial College London
Johannes Totz   Imperial College London
Maryam Rettmann   Mayo Clinic
Kirby Vosburgh   Harvard University
Kenko Fujii   Imperial College London
Simon Drouin   McGill University
Danail Stoyanov   Imperial College London
Tamas Ungi   Queen's University
George Stetten   Carnegie Mellon University
Jue Wu   University of Pennsylvania

Table of Contents

The Patient Comes First: The Perfect Surrogate

Richard A. Robb

Biomedical Imaging Resource, Mayo Clinic,
Rochester MN USA
`robb.richard@mayo.edu`

The information age has been a marvel to behold. Mind boggling advances in almost every sphere of science and technology have been forthcoming - discoveries, creations, processes, devices, instruments, systems that are smaller, bigger, faster, and better. This accelerated progress has also been manifest in medicine where significant advances in the ability to more effectively diagnose, treat and cure have gone forward at a remarkable pace. This also holds true for most sub-disciplines within medicine, including the focus of this workshop on augmented environments and computer aided interventions (AE-CAI). These are beginning to have measurable impact on the health and well-being of the population. Initial technologies have exhibited high performance and realistic simulation for efficient training, rehearsal and on-line guidance of medical interventional procedures. But this success still lags far behind the ultimate potential, and even farther from the horizon of expectations and vision for the field.

There are two often overlooked and/or marginalized but ever present hurdles: 1) putting the patient first (or not) to target and focus on patient relevant advances, and 2) effective, expedient translation from basic research and engineering prototypes to clinical products. The era of individualized medicine has begun — the so-called P4 medicine — pre-emptive, predictive, personalized and participatory. The respective fronts of progress toward fully realizing these four inter-related goals vary in pace of progress and delivery of expected results, in spite of the intrinsic advantages offered by modern technology. For example, high spatial and temporal resolution imaging of the entire body provides not only a patient-specific dataset to use, but also a faithful surrogate of the patient to facilitate development and application of an individualized, targeted, focused, validated prototype, and, finally a clinically accepted product. We are not yet taking full advantage of the (almost) perfect patient surrogate available to us for rapidly and accurately translating technological advances into human clinical applications.

This volume will put some emphasis on imaging science and technology engineering as related to augmented environments for computer-assisted interventions, and towards effective production and translation of working, validated prototypes and their integration into relevant healthcare workflows and solutions focused on the individual patient. A patient-first driven process influences design and development of optimized clinical workflows to realize the objective of serving the individual patient best. Putting the patient first insures that whatever the pathway to success may be, the objective and solution have positive impact on the well-being of the patient. The planning for the translational

system incorporates upfront the ideal patient surrogate (e.g., as facilitated by high resolution faithful imaging) and is used to bring specific attention and information to solving the patient's problem. Indeed, the over-arching solution covers the P4 space of both current future medicine — pre-emptive (prevention), predictive (estimation of risk/benefits), personalized (designed for the individual patient), and participatory (physicians and surgeons working closely with scientists and engineers keeping the patient first in the overall design and workflow integration of the targeted solution).

The opening presentation illustrated several real clinical individual "patient-first" examples of successful AE-CAI approaches/systems wherein success was obviously produced by using an ideal patient surrogate — a faithful specific full 3D or 4D image of the patient. Hopefully these examples may help inspire accelerated efforts toward achieving the long promised potential of AE-CAI that remains too marginalized and too hidden away in a small corner of the ever expanding fortress of the information age.

Biography: Dr. Richard A. Robb received the B.A. degree in Mathematics in 1965, the M.S. degree in Computer Science in 1968, and the Ph.D. degree in Electrical Engineering (Computer Science) in 1971, all from the University of Utah. He is Professor of Biophysics and Professor of Computer Science in Mayo Clinic College of Medicine, Medical School, Mayo Graduate School and the Department of Physiology and Biomedical Engineering at Mayo Clinic. In 2000 he was awarded a Mayo named professorship — The Scheller Professorship in Medical Research — that endows his research and educational activities, citing him as a "respected leader, innovative scientist, and caring educator". In 2011 he received the Distinguished Alumnus Award from the University of Utah. Dr. Robb is Director of the Biomedical Imaging Resource Lab at Mayo Clinic. His research interests are in development of modular image analysis software systems for advanced diagnosis and treatment of disease, and in development, validation and deployment of real-time minimally invasive image-guided systems for medical interventions. He has mentored 25 Ph.D. Graduates, 20 Post-doctoral Research Fellows and over 100 visiting students and trainees. Dr. Robb has over 450 publications in the field, including 5 books and 30 book chapters and holds several patents related to inventions in multi-dimensional image display, manipulation and analysis. Dr. Robb has had a productive career of over 40 years in biomedical imaging science research, validation and translation to clinical utility and graduate biomedical education. His research has been funded over that time primarily by NIH grants, but also by philanthropic foundations and industrial sponsors. These systems have been successfully translated to specific clinical uses in brain cancer, bone cancer, lung cancer, colon cancer, prostate cancer, epilepsy, osteoporosis, heart and coronary artery disease, emphysema, and craniofacial reconstruction, among other applications. Software systems developed in his laboratory have been licensed to several companies and disseminated to and are being actively used in over 800 institutions around the world.

Time to Go Augmented in Vascular Interventional Neuroradiology?

René Anxionnat[1], Marie-Odile Berger[2,3], and Erwan Kerrien[2,3]

[1] Therapeutic and Diagnostic Interventional Neuroradiology Dept, Univ. Hospital, Nancy, F-54000, France
[2] Inria, Villers-lès-Nancy, F-54600, France
[3] Université de Lorraine, Loria, UMR7503, Vandœuvre-lès-Nancy, F-54600, France

This editorial paper reports on our experience in introducing augmented reality (AR) in interventional neuroradiology environments. Our expectations about the next AR tools, in particular for more advanced visualization, are also put forward. For practical reasons, the references will be restricted to our contributions. For further information, the last recommendations concerning the medical management of aneurysm induced hemorrhages can be found in [1].

1 Interventional Neuroradiology

Interventional neuroradiology is a medical discipline that leverages minimally invasive techniques to operate on vascular lesions of the nervous system. The treatment is performed through the inside of the blood vessels, using medical imaging as visual feed-back. Such endovascular procedures are mainly performed to treat aneurysms. Intracranial aneurysms are sacciform ectasias of arteries located on the brain surface. The main associated risk is their rupture, leading to an hemorrhage with dramatic clinical consequences. Their size (diameter) varies between 2 mm and a few tens of millimeters. An endovascular treatment (EVT) progressively occludes the aneurysm with very soft spiral-like coils. EVTs were introduced in the 60's but really took off in the 90's after controlled detachable coils were developed. EVT has now replaced neurosurgery as the first choice treatment of intracranial aneurysms.

EVTs are performed in interventional vascular operating rooms where X-ray imaging enables the precise analysis of the lesions and their treatment. X-ray Angiography still remains the imaging modality of reference thanks to its optimal spatial and temporal resolution. The endovascular navigation followed by the EVT itself are performed under the real-time visual feed-back of a low dose and high frame rate (15 fps on average) angiography image acquisition called *fluoroscopy*. However, these low quality images are projection images that do not contain any 3D information. Several viewpoints are required for an easier guidance through complex vascular networks and a better control of the treatment. Therefore aneurysm EVT are usually performed in biplane mode, providing simultaneous Antero-Posterior (AP) and Lateral (Lat) fluoroscopic views. No significant improvements were brought in the last few years in the control of

EVTs using fluoroscopy. The development of Augmented Fluoroscopy very recently modified this status by introducing, for the first time, Augmented Reality in vascular interventional neuroradiology.

2   Augmented Fluoroscopy: Principle and Clinical Interest

The idea of enriching fluoroscopy images with 3D data was born shortly after the development, in 1997, of 3D rotational angiography (3DRA) [2], a cone-beam tomography reconstruction techniques that produces a volume from angiography images acquired during a rotation of the vascular C-arm. 3DRA allows for an optimal treatment planning thanks to a precise analysis of the aneurysm morphology: its size, its shape and its neck, where it connects with its parent vessel [3]. From 3DRA, a working view can be defined that best identifies the aneurysmal neck and clearly reveals the limits for the coil positioning. 3DRA rapidly became a must-have imaging modality for EVT planning, to improve treatment safety and expand their indications. However, over a decade was necessary to actually integrate 3DRA with fluoroscopy thereby building the first augmented reality tool suitable for use in clinical practice.

During EVT, superimposing 3DRA onto the fluoroscopy sequence provides the physician with a stereoscopic vision of the morphology of vascular bifurcations, of the aneurysm and its relations with neighboring vessels. Endovascular guidance is made easier and coil deployment is better controlled.

Augmented Fluoroscopy (AF) is based on 3D/2D registration between 3DRA and fluoroscopy in any C-arm orientation. A prototype was developed in close collaboration between the research team (Inria Nancy Grand-Est), the manufacturer (GE Healthcare) and the medical staff (Department of Neuroradiology University Hospital of Nancy, France). The geometric accuracy of the implemented machine-based registration was first measured [4]. The next step consisted in further evaluating the prototype in the clinical conditions of an aneurysm EVT. Validation criteria were designed to address the clinical requirements met in the operating room, in terms of geometric accuracy and visualization. Visualization should indeed both preserve the visibility of the EVT tools (microcatheters, guidewires, and coils) and clearly depict the vascular topology in 3D.

The geometric accuracy was first estimated on a silicon phantom of the cerebral vasculature under various orientations of the C-arm. It was proved to be less than 0.5 mm and therefore compatible with the clinical objective [5]. This first conclusion was confirmed on clinical cases.

Concerning visualization, we chose to blend the 3DRA volume in Volume Rendering onto the fluoroscopic images. Easy to implement, this solution presents as a further advantage that it barely impacts the usual physician's environment (see Figure 1). Using AF in clinical conditions confirmed the relevance of this choice, and helped to significantly improve this prototype with functionalities such as interactive tuning of the vessel transparency to reach an optimal consensus between vascular relief and envovascular material visibility. After these successive improvements, AF was launched as a product in 2007.
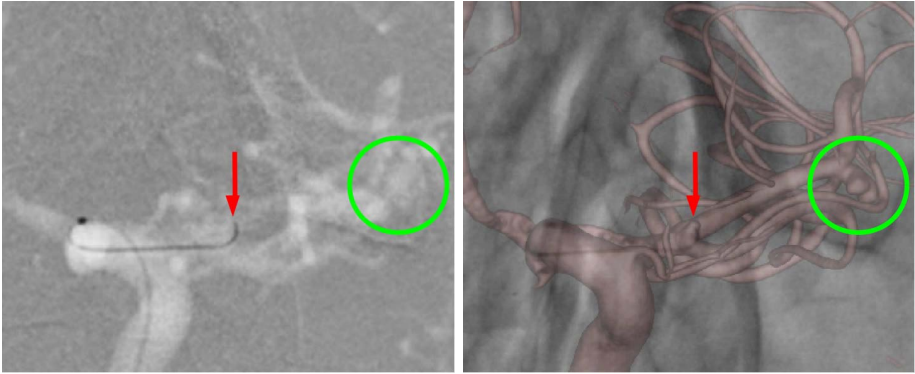
Fig. 1. Augmented Fluoroscopy (AF) case. (left) Original fluoroscopy image ; (right) AF view. AF offers a much clearer depiction of the ruptured aneurysm to be treated (green circle), and also unveils a microscopic incidentally discovered aneurysm which hampers the guide wire navigation (red arrow).

3 The Ordeal of Going into Clinical Practice: An Under-Estimated Step

The medical team was involved in the development of AF right from the start, and the preliminary evaluation demonstrated its definite clinical interest. Nevertheless, several limitations made the integration of AF in clinical routine very gradual. First, AF is only available on the AP C-arm. Yet, in biplane mode, the working view used for the EVT may only be reached by the Lat C-arm. Second, in actual clinical practice, many events may cause the machine-based registration to degrade which quite often calls for a manual correction in the absence of an image-based registration adjustment.

Moreover, the AF monitor was first set aside in the operating room in order to minimize the perturbation on the otherwise unchanged environment. On one hand, the physicians readily accepted such a discrete modification in their work environment. On the other hand, the prototype was underutilized. Therefore, significant improvements, in terms of ergonomics, were made to better integrate the prototype into the operating room, such as larger zoom factors and an actual integration as part of the panel of control monitors.

The second step towards the use in clinical routine was gradual and more or less fast depending on physicians. Indeed, physicians needed some time to agree to change their work routine either during risky interventions that required a lot of concentration or for cases they could easily treat without AF. But this phase, including, and especially for the latter simple cases, proved essential to the evaluation by physicians who had to become familiar with the new environment and compare it with reference tools, namely casual fluoroscopy that was still available. Once AF was mastered, the physician could use it in complex cases

and rely on information that would not be accessible with mere fluoroscopy, e.g. axial views of complex aneurysms. AF is being used in clinical routine at our institution since 2011.

4   Perspectives: Other Applications of Augmented Fluoroscopy; Needs for Further Augmented Reality Tools

AF marks a breakthrough in intraoperative visualization by introducing the third dimension. It was perfectly integrated into the EVT environment for intracranial aneurysms. Its use for other clinical applications and in combination with other 3D imaging modalities were logically pursued.

Arteriovenous malformations (AVMs) are scarcer than brain aneurysms and also differ a lot on both the morphology and therapy levels. They consist of a network of abnormal vessels interposed between arteries and veins which creates a high blood speed arteriovenous shunt in the brain parenchyma. The treatment is to occlude the AVM using fluid or pasty injections after a distal navigation within a very small, sub-millimetric, highly branched vascular network. Technological advances in the field of navigation tools and endovascular material allow to perform such embolization in very distal vascular areas, as well as highly functional brain regions.

AF can facilitate this complex navigation by displaying the origin and path of afferent arteries and differentiating them from neighboring vessels or vessels that superimpose in the fluoroscopic view. These clinical needs are more demanding than for aneurysms given the complexity and finesse of the vascular network. Additional work in terms of accuracy of registration, and vascular segmentation and visualization will be necessary to make this tool functional in clinical practice.

During an EVT, the physician can only see vessels on the monitors, but has no access to a direct visualization of the cerebral parenchyma in the surroundings of the endovascular material. Locating crucial functional brain areas cannot be realized but through a mental effort to relate a priori anatomical knowledge and preoperative imaging to intraoperative views. The inherent inaccuracy of such process may lead the physician to excessively embolize in highly functional areas, potentially causing dramatic complications, or, conversely to give up treatment in case of doubt. A multimodal localization is definitely required, especially combining MRI and angiography, to identify the cerebral regions with a high accuracy. Innovative visualization tools will have to be developed to setup such new multimodality environments and bring relevant information, in an intuitive way, to the physician during the treatment.

During the treatement, AF provides 3D vascular information in any orientation the C-arm can reach. As such, a real progress was made compared to mere fluoroscopy imaging. However, endovascular material still is only visible under projection in limited orientations. The next step will see the 3D reconstruction of endovascular tools and material during the EVT. These tools and material

will be directly located in the 3DRA volume and thereby, on one hand, visible in any view, including virtual views, such as axial views inaccessible to the C-arm, and on the other hand, easy to relate to any 3D multimodal imaging volume. Very hard technical and scientific issues will have to be overcome, but it casts a glimpse of potentially dramatic breakthroughs in the way EVTs are performed.

Having such a possibility will allow for a better control of the EVT, thus improving safety. However working in a multimodal environment will generate inaccuracies due to the involved registration and segmentation processes, that will need to be quantified and that the physician should be fully aware of. The physician will have to agree to cope with this uncertainty when working on virtual views without the direct control provided by fluoroscopy, and thus no possibility for correction. An important intellectual investment will be necessary to learn how to use these tools and manage this uncertainty before appropriating these tools in clinical practice.

Eventually, future augmented environments will be enriched not only with visual information from other imaging modalities such as functional, but also with other data such as generated by hemodynamic simulation or wall pressure computations for aneurysms.

5   Conclusion

Our experience with AF shows it is possible to develop Augmented Reality tools in the field of interventional neuroradiology. But there is a long way between the initial idea and the product used in clinical practice: 10 years in the case of AF.

For this tool to be used it must meet specific clinical needs, clearly identified by physicians, and be validated for this purpose. This validation must meet specific criteria and objectives. In the case of an EVT, which corresponds to a risky procedure requiring a high concentration from the physician, any Augmented Reality tool must be fully integrated into the operating room and provide an intuitive interface.

It is important that any operator that was not necessarily involved in the development and validation process, be persuaded to use the tool first in simple cases where it can be evaluated within a personal operating environment so as to gradually become more familiar with the tool and finally trust the additional information it provides. Under these conditions the augmented environment has the potential to bring real benefits to the physician, especially in complex cases, and imply actual changes in the clinical practice. As an example, AF provides actual 3D information which biplane fluoroscopy useless for most interventions and thus reducing patient irradiation and exam time while probably improving the safety of the intervention.

AF has paved the way for Augmented Reality in interventional neuroradiology. A vast field of opportunities exist to adapt this tool to other, more demanding, clinical applications and develop new Augmented Reality tools. New angiographic rooms will emerge that integrate an augmented environment, allowing the interventional neuroradiologist to work in fully multimodal setup enriched with visual, and other kinds of, information.

References

1. Connolly Jr., E.S., Rabinstein, A., Carhuapoma, J.R., Derdeyn, C.P., Dion, J., Higashida, R., Hoh, B., Kirkness, C., Naidech, A., Ogilvy, C., Patel, A., Thompson, B., Vespa, P.: Guidelines for the management of aneurysmal subarachnoid hemorrhage. Stroke 43, 1711–1737 (2012), doi:10.1161/STR.0b013e3182587839
2. Anxionnat, R., Bracard, S., Macho, J., da Costa, E., Vaillant, R., Launay, L., Trousset, Y., Romas, R., Picard, L.: 3D angiography. Clinical interest. First applications in interventional neuroradiology. J. Neuroradiol. 25, 251–262 (1998)
3. Anxionnat, R., Bracard, S., Ducrocq, X., Trousset, Y., Launay, L., Kerrien, E., Braun, M., Vaillant, R., Scomazzoni, F., Lebedinsky, A., Picard, L.: Intracranial aneurysms: Clinical value of 3D digital subtraction angiography in the therapeutic decision and endovascular treatment. Radiology 218, 799–808 (2001)
4. Gorges, S., Kerrien, E., Berger, M.-O., Trousset, Y., Pescatore, J., Anxionnat, R., Picard, L.: Model of a vascular C-arm for 3D augmented fluoroscopy in interventional radiology. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 214–222. Springer, Heidelberg (2005)
5. Gorges, S., Kerrien, E., Berger, M.O., Pescatore, J., Trousset, Y., Anxionnat, R., Bracard, S., Picard, L.: 3D augmented fluoroscopy in interventional neuroradiology: precision assessment and first evaluation in clinical cases. In: Proc. of AMI-ARCS 2006 (2006)

Endoscopic Image Overlay for the Targeting of Hidden Anatomy in Laparoscopic Visceral Surgery

Matteo Fusaglia[1], Kate Gavaghan[1], Guido Beldi[2], Francesco Volonté[3],
François Pugin[3], Matthias Peterhans[1], Nicolas Buchs[3], and Stefan Weber[1]

[1] ARTORG Centre for Biomedical Engineering Research, University of Bern,
Bern CH-3010, Switzerland
{matteo.fusaglia,kate.gavaghan,matthias.peterhans,
stefan.weber}@artorg.unibe.ch
[2] Department for Visceral Surgery and Medicine, Bern University Hospital,
Bern CH-3010, Switzerland
guido.beldi@insel.ch
[3] Visceral and Transplantation Surgery, Department of Surgery, University Hospitals of
Geneva, Rue Gabrielle-Perret-Gentil 4, CH-1211 Geneva 14, Switzerland
{francesco.volonte,francois.pugin,nicolas.c.buchs}@hcuge.ch

Abstract. Limitations associated with the visual information provided to surgeons during laparoscopic surgery increases the difficulty of procedures and thus, reduces clinical indications and increases training time. This work presents a novel augmented reality visualization approach that aims to improve visual data supplied for the targeting of non visible anatomical structures in laparoscopic visceral surgery. The approach aims to facilitate the localisation of hidden structures with minimal damage to surrounding structures and with minimal training requirements. The proposed augmented reality visualization approach incorporates endoscopic images overlaid with virtual 3D models of underlying critical structures in addition to targeting and depth information pertaining to targeted structures. Image overlay was achieved through the implementation of camera calibration techniques and integration of the optically tracked endoscope into an existing image guidance system for liver surgery. The approach was validated in accuracy, clinical integration and targeting experiments. Accuracy of the overlay was found to have a mean value of 3.5 mm ± 1.9 mm and 92.7% of targets within a liver phantom were successfully located laparoscopically by non trained subjects using the approach.

Keywords: Augmented reality, Laparoscopic surgery, Endoscopic image overlay.

1 Introduction

The use of laparoscopic techniques in abdominal surgery eliminates the need for large incisions across the abdomen (such as the 24 to 30 inch chevron incision) and thus, reduces postoperative discomfort, length of the operation, hospital stay time and recovery time in addition to complications associated with such invasive procedures [1][2] .

Traditionally in laparoscopic surgery, visual information pertaining to the surgical situation, including the laparoscopic instruments within the abdomen, is fed back to the surgeon in the form of 2D endoscopic video images on a standard monitor.

This technique results in a loss of three-dimensional (3D) viewing and a limited field of view of the surgical situs. Such deficiencies weaken spatial understanding and depth perception and increase the risk of undesired contact between instruments and anatomical structures, with potential damage to the patient. Furthermore, endoscopic images provide visualization of surface structures only. Hidden targets must be located in preoperative images or intra-operative ultrasound images and mentally aligned to the endoscopic view. For these reasons, the use of endoscopic 2D visual guidance requires significant hand-eye coordination and anatomical understanding that results in a need for highly trained surgeons.

In an attempt to enhance the presented visual guidance information, and thus improve intuitiveness of laparoscopic surgery, a number of alternative visualisation approaches based on augmented reality (AR) techniques have been investigated by various groups. In [3],[4],[5] Marescaux, Nicolau and Soler et al present an augmented reality system for liver and pancreatic surgery which overlays 3D internal structures onto the endoscopic view. Additionally, Feuerstein et al presented a CT based image overlay system for the planning of port placement in abdominal laparoscopic surgery. Whilst the systems were tested in isolated clinical cases or animal studies for feasibility, depth perception issues were not addressed and the usefulness of the systems in targeting underlying structures have not been evaluated. Works by Hong et al [6] and Winne et al [7] explored the use of image overlay AR systems with varying depth indicators for endoscopic Oto-Rhino-Laryngologie applications with varying success but, no translation to abdominal surgery has been reported.

An exhaustive review of previously developed AR systems for laparoscopic surgical oncology is presented by Nicolau et al in [8].

More recently, technological advances in stereoscopic endoscopy have enhanced depth perception by allowing surgeons to view the surgical scene in 3D. However, associated high costs and the need for special eye wear have inhibited the acceptance of available systems. Additionally, such technology does not attempt to solve issues pertaining to the limited field of view or to the view of underlying structures.

In this work, we present an endoscopic image overlay AR system for abdominal laparoscopic surgery that aims to assist in the accurate location of underlying structures and the reduction of damage to surrounding critical structures. The system allows underlying anatomical 3D models to be viewed in the same view as endoscopic images. Additionally, we attempt to address the lack of depth perception by adding a novel feature which depicts spatial and distance information related to specified hidden targets.

An evaluation of the systems accuracy and clinical integration and a quantitative analysis into its usefulness in targeting underlying structures is presented.

2   Methods

2.1   Endoscopic Image Overlay

System Overview

An instrument guidance system for open liver surgery (CAScination, CH) [9], was augmented with laparoscopic tool navigation capabilities and a software module to interface with a standard endoscopic device and enable AR endoscopic overlay visualisation.

The tracking of endoscopic instruments is realized via the attachment of retro-reflective markers that are tracked by the navigation system's passive optical tracking system (NDI Vicra, Northern Digital, Canada).

Virtual 3D models, as for conventional image guided liver surgery, are segmented from pre-operative medical images (i.e. CT) (meVis distant Services, Germany) [10]. Through the use of a tracked palpator, anatomical models can be laparoscopically registered to the patient, via a locally-rigid landmarked-based registration [9]. Thereafter, models are displayed in a virtual scene relative to tracked surgical tools, in the tracker coordinate system by applying the registration transformation, $^{tracker}T_{patient}$.

AR is provided by overlaying, in real-time, a virtual scene of the patient, captured at the calibrated view of the endoscope, onto the endoscope's video stream. By rendering the virtual scene of the patient at the same view as the endoscope, the need for real time image registration is avoided.

Endoscope Calibration

In order to overlay patient data onto endoscope video stream, the rigid transformation relating the coordinate system of the tracker attached to the endoscope to the coordinate system of the endoscope camera, $^{endoscope}T_{camera}$, is required.
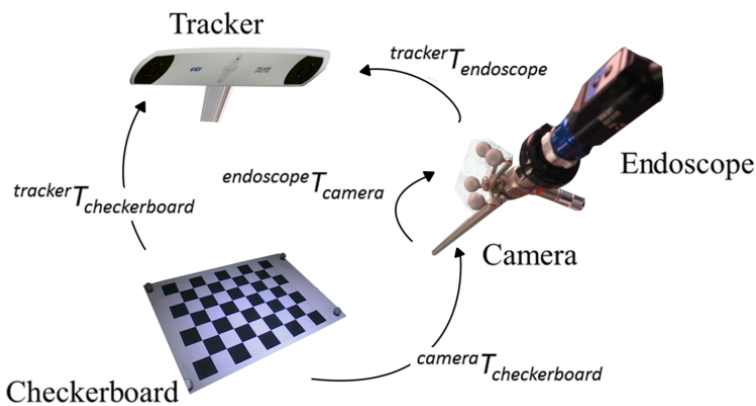
Fig. 1. Endoscope calibration transformation diagram

An automatic calibration process, based on camera calibration techniques and the methodology presented by Gavaghan et al [11], was developed in order to determine the required transformation according to Eq. 1.

$$\left(^{endoscope}T_{camera}\right) = \left(^{tracker}T_{endoscope}\right)^{-1} \cdot \left(^{tracker}T_{checkerboard}\right) \cdot \left(^{camera}T_{checkerboard}\right)^{-1} \quad (1)$$

The calibration process utilizes a calibration pattern (i.e. checkerboard), composed of a rectangular grid (9 x 7 black-white pattern) attached to a metal plate together with four passive markers. With the endoscope in a static position, *n* different images of the checkerboard at different angles and positions are acquired, together with the endoscope position, $^{tracker}T_{endoscope}$, and the checkerboard position with respect to the tracking device coordinate system, $^{tracker}T_{checkerboard}$ (refer to Fig.1).

The transformation $^{camera}T_{checkerboard}$ and the intrinsic camera parameters of the endoscopic camera, were determined by solving the relationship between 3D checkerboard corner positions and the corresponding 2D image pixel location using a Zhang based [12] calibration methodology and the OpenCV camera calibration library.

The transformation, $^{camera}T_{checkerboard}$ is given by the calculated extrinsic parameters. For each set of extrinsic parameters, the transformation $^{endoscope}T_{camera}$ is obtained by solving Eq. 1. Error introduced during the calibration phase yields variance across the transformations. The transformation which minimises re-projection error was selected [11].

Endoscopic Image Overlay Functionality
After patient registration and endoscopic calibration, the virtual scene can be rendered at the endoscopic view (refer to Fig. 2) and overlaid on the endoscopic image.
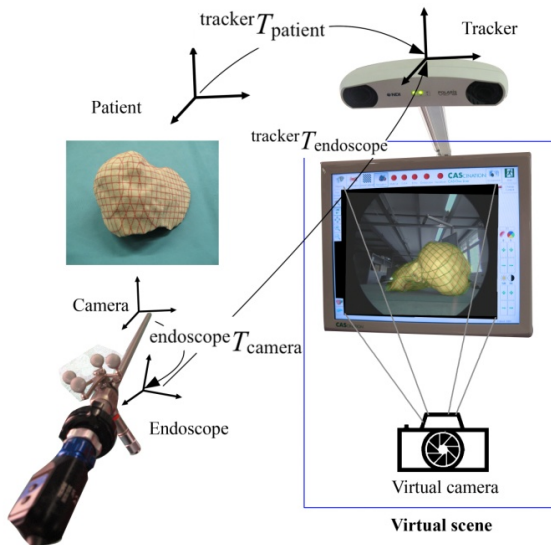
Fig. 2. Endoscopic image overlay functional model

The endoscopic view is captured via the definition of a virtual camera (Open Inventor^{TM}) within the virtual scene by the intrinsic parameters of the endoscope according to the relationships presented by Li et al [13]. The pose of the camera is defined in real time using the tracked position of the endoscope, $^{tracker}T_{endoscope}$, and the calibration transformation according to Eq. 2.

$$\left(^{tracker}T_{camera}\right) = \left(^{tracker}T_{endoscope}\right) \cdot \left(^{endoscope}T_{camera}\right) \tag{2}$$

Endoscopic Image Augmentation

To enable the visualization of underlying structures, 3D anatomical models of critical anatomy along with a virtual model of the current tool location can be overlaid with the endoscopic image. In order to cope with the loss of depth perception related to endoscopic images and the overlay of virtual anatomical models, additional guidance of a tool to a target was integrated. The guidance information, based on the application described by Gavaghan et al [14] consists of a projected trajectory line overlaid on the endoscopic image, a cross-hair target and a depth-bar, which depicts the distance from the tip of the tool, and the target itself (Fig. 3). Internal 3D anatomical structures (e.g. tumours) can be defined as targets and thereafter become active within the targeting guidance application.

By aiming the tracked tool approximately at the target, the target model is visualised within the cross hair target and the current distance to the target is displayed on the depth bar. By visualizing the projected line from the tool to the target, the surgeon can assess the position of critical structures and adjust their approach accordingly. By adjusting the angle of the tool such that the projected line points to the target, the tumor will be brought into the center of the cross hair. The depth bar appears full once the target is reached and becomes red if the required depth is exceeded.

Fig. 3. Augmented endoscopic visualization with anatomical structures and guidance data

2.2   Evaluation

In order to validate the proposed augmented reality visualization approach three experiments were conducted. Firstly, the accuracy of the overlay was determined on a 3D anatomical phantom, secondly 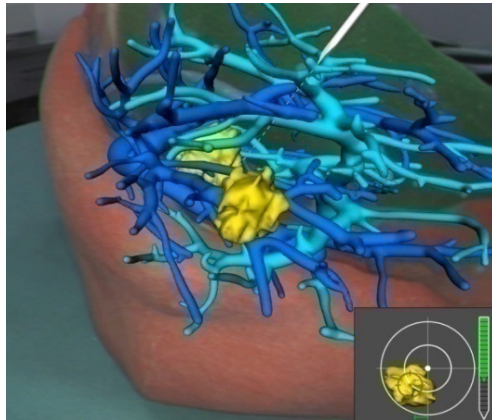the feasibility and clinical integration of the approach was verified interpretatively and finally, the usefulness of the augmented endoscopic view in guiding the user to underlying invisible landmarks was quantitatively assessed.

Accuracy Evaluation

The overall accuracy of the augmented reality system was evaluated using a rapid prototyped model of a human liver with a superimposed 1-cm surface grid (Fig. 7). After calibration of the endoscope camera, AR images were created by superimposing the endoscope image with a 3D image of the model (Fig. 7), using three different orientations of the endoscope with respect to the model, and two different distances between the endoscope camera and the model (approximately 2 cm and 35 cm). From each AR image, discrepancies between the grid in the endoscope image and the 3D image were measured on 8 different nodes: 4 in the center of the image, and 4 close to the outer edge of the liver model.

Clinical Integration

The workflow and clinical integration of the augmented reality system was tested in a single case of laparoscopic abdominal tumor resection. The surgeon's endoscope was calibrated using the above described semi-automatic process prior to surgery. Thereafter, the navigation system was positioned on the opposite side of the patient to the operating surgeon with the tracking camera positioned above the patient. The endoscope was draped and a sterilised tracking marker was attached on top of the sterile drape (refer to Fig. 8). The patient was registered by matching four anatomical landmarks (three internally selected landmarks and one landmark located on the sternum). 3D models of a tumor, spleen and vessels, segmented from CT images, were overlaid onto the endoscope's stream during the procedure. Augmented endoscopic views were used to aid in invisible structure localisation.

Laparoscopic Targeting of Underlying Structures

The usefulness and effectiveness of the described image overlay technique in assisting in the location of non-visible internal targets, was quantitatively assessed in phantom experiments.

A virtual model with hepatic vein, portal vein and liver surface was segmented from patient CT data by MeVis distant services, Germany. A portion of the anterior liver surface was removed and the remaining surface was physically constructed using rapid prototyping 3D printing. The rigid phantom shell was lined internally with a sterile drape and filled with porcine liver tissue (refer to Fig. 4).

Eleven spherical landmarks with 2 cm diameter were added to the virtual liver model in varying complex positions, avoiding the vessel branches (refer to Fig. 4). The phantom was secured within a laparoscopic surgical trainer (Body Torso Laparoscopic Trainer, Pharmabotics Ltd, UK).
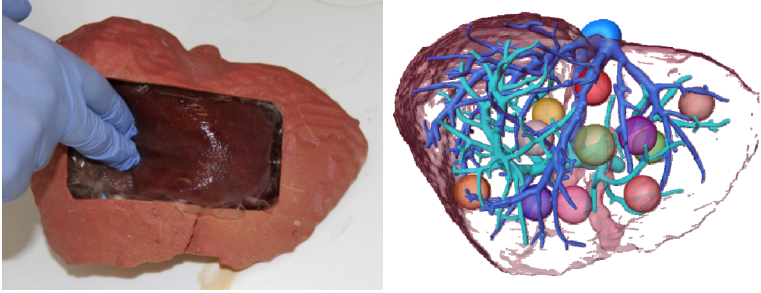
Fig. 4. Rigid liver phantom with porcine liver tissue (left) and virtual liver model with hepatic and portal veins, and 11 spherical targets (right)

As the overlay was aimed at improving intuitiveness and ease of laparoscopic targeting procedures, five subjects, untrained in laparoscopic or other surgical techniques were selected for testing. All users were familiar with the theory of image guided surgical techniques and systems. Subjects were required to laparoscopically locate all internal landmarks in a defined order with a tracked ablation needle (Cool-tip[TM] RF ablation system, Covidien, Ireland) using the described augmented endoscopic view. Subjects were given five minutes training period immediately prior to testing.
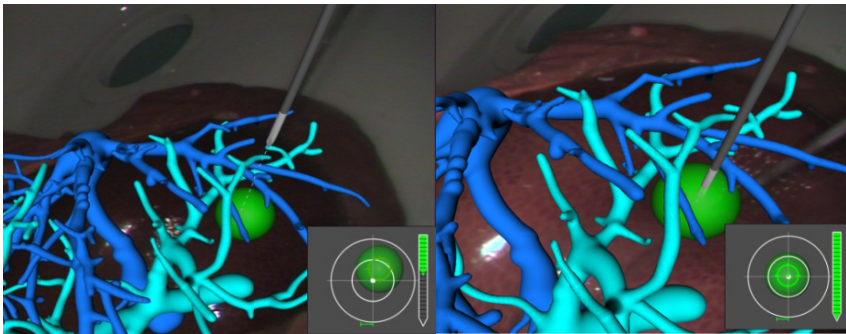
Fig. 5. Augmented endoscopic view as displayed during landmark targeting with vessels, trajectory projection, target information and depth bar

Targeted landmarks were displayed one at a time in conjunction with vessel structures and targeting and depth information within the augmented endoscopic view (refer to Fig. 5) on a nearby monitor (refer to Fig. 6).
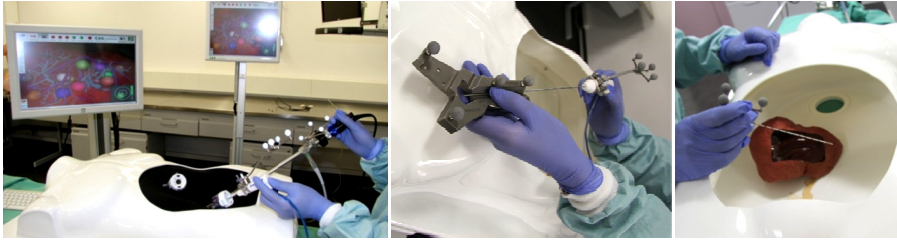
Fig. 6. Laparoscopic testing setup with augmented endoscopic view (left); Needle calibration (middle); liver registration within the laparoscopic trainer (right)

Users were instructed to position the tip of the needle within the target without damaging surrounding vessel structures. Users were advised that the damage of peripheral vessels was less critical than damage to main vessel branches and that the time taken to locate the target would be recorded. The endoscope was operated by the user or by an assistant as per the users preference. Prior to each test, the ablation needle and registration pointer were calibrated and the liver rigidly registered to the virtual model via the processes described by Peterhans et al [9] (Fig.6). The position of the optically tracked ablation needle was recorded throughout the targeting of each landmark and the targeted position (as decided by the user) was recorded by the optical tracking system. The tracked needle path and target positions were later plotted with 3D anatomical models in Amira® (Visage Imaging, USA). Target positions located within the volume of the target were defined as successfully reached. Distances between non reached targets and the spherical landmarks were determined. Additionally, crossings of the traced path with vessel models were visually determined and recorded and categorized as either main vessel damage or peripheral vessel damage depending on where the crossing occurred.

3   Results

Accuracy Evaluation
Accuracy evaluation studies on anatomical phantoms (Fig. 7) showed that the endoscope's stream could be overlaid with 3D images with a mean error of 3.5 mm ± 1.9 mm. The error comprises calibration and instrument tracking noise, as well as rigid registration error. The accuracy achieved is suitable for clinical use in laparoscopic liver surgeries.
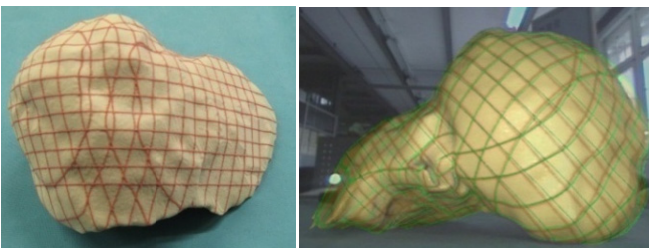
Fig. 7. Liver testing phantom with grid (left) and AR image (right)

Clinical Integration

During the application of the approach within a clinical scenario, workflow and integratability were validated. A tracked marker was fixed to the surgeon's endoscope which could thereafter be tracked by the navigation systems optical tracking camera. The endoscope was calibrated prior to the sterile phase in approximately 5 minutes.
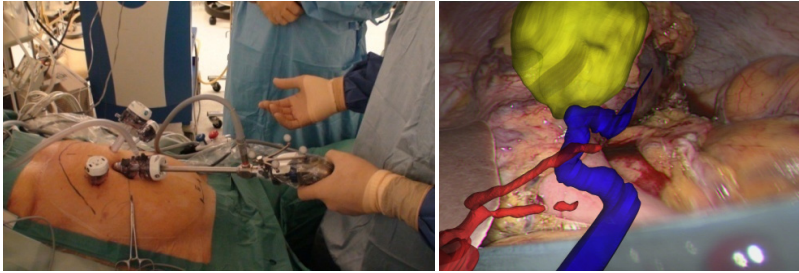
Fig. 8. The tracked endoscope employed during laparoscopic tumor resection surgery (left) and the augmented endoscopic display with tumor and surrounding vessels (right)

Overlay of 3D data (i.e. vessels, tumor, organs) helped in locating underlying and critical structures and did not extensively inhibit the view of endoscopic images (refer to Fig. 8).

Hidden Structure Targeting

Of N = 55 targeted landmarks, 51 were successfully reached. Three of the five subjects targeted all 11 landmarks successfully.

The remaining four targets were missed by two of the subjects by distances of 1.87 mm, 2.4 mm, 5.69 mm and 25.95 mm respectively. 52.7% of the targets were successfully reached without damaging (crossing) surrounding vascular structures.
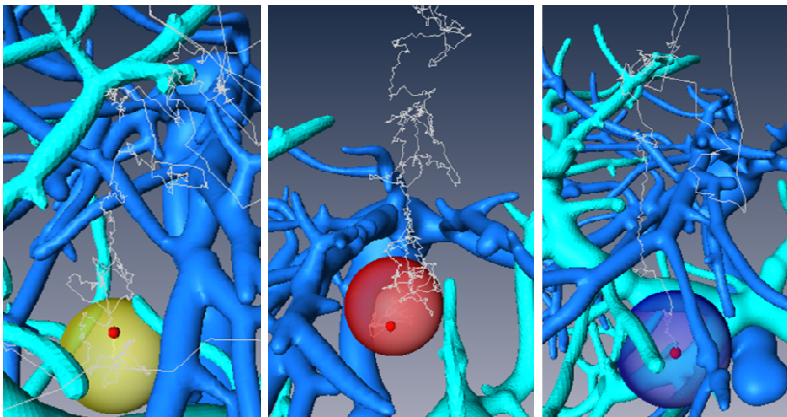
Fig. 9. Tool path (white) and target position (red) during the successful location of landmarks without vessel damage
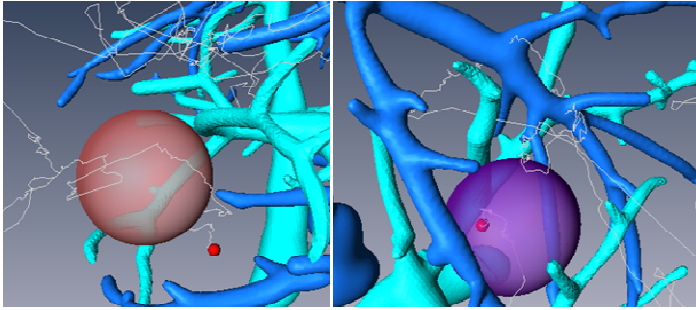
Fig. 10. A missed target (left) and damage to a primary vessel branch (right)

Images of the tool path and final target position for three such successful cases are displayed in Fig. 9. Primary vessel damage (crossing of the main vessel branches by the tool) occurred in 18% of cases. Peripheral vessels were damaged in 34% of cases. The mean time required to reach a target was 91 seconds. Results of the experiment are summarized in Table 1. Plots demonstrating a missed target and primary vessel damage are displayed in Fig. 10.

Table 1. Hidden landmark targeting results summary

N	Targets reached	Targets reached without vessel damage	Primary vessel damage	Peripheral vessel damage	Mean time
55	92.7 %	55 %	18 %	34 %	91 s

4   Discussion

Within this work, we have presented an augmented endoscopic visualization approach for laparoscopic surgery that aims to: improve depth perception, provide additional tool positioning information, improve intuitiveness, and allow the viewing of underlying hidden structures. Integration of the approach into an existing surgical navigation system for surgery allowed the system's inherent verified patient registration and tracking capabilities to be utilised. Furthermore, minimal additional equipment is required, ensuring that costs and obtrusiveness are minimized. Through calibration and tracking of the endoscope, the need for computationally costly image registration is eliminated, allowing overlay to be displayed in real time (>20Hz).

The use of optical tracking requires line of sight, and thus, somewhat limits the orientation of the endoscope. However, in an initial clinical application such deficiency could be coped for by the attending surgeon during times at which the augmented view was required. During targeting evaluations in which a 0 degree scope was used, the subject's view and mobility were only minimally inhibited by the need for line of sight. Whilst the use of non optically based tracking systems would

remove the line of sight requirement, any usability improvement would need to be weighed against possible inferior accuracy associated with electromagnetic or other tracking modalities [15][16] .

Whilst rigid registration techniques employed within these evaluations cannot cope for movement and deformation of an entire organ, the techniques are currently used in commercial navigation systems for soft tissue surgery and have been shown to be sufficient for small areas of interest (i.e. area surrounding a tumor) within deformable anatomy [9] [17]. Additionally, re-registration can be performed intermittently using pair point matching or ultrasound based registration techniques [18].

Perception of depth remains a great challenge for endoscopic image overlay. Within the presented system, the perception of depth of specified target structures has been improved through the overlay of additional depth and trajectory information. The depth bar and cross hair improve depth perception and reduce hand eye coordination complexities by decomposing guidance data into one and two dimensions respectively. Depth perception of surrounding structures was not yet addressed and the effect of this was evident in the resulting higher percentage of vessel damage than target misses. Methods of encoding depth in anatomical models have been investigated by others [19] and such techniques may be integrated into the overlaid virtual models in future work.

The ability of non trained subjects to target relatively small invisible structures with a success rate of 92.7 % demonstrated that the described augmented endoscopic view could aid in the improvement of targeting accuracies and could possibly result in reduced procedure time and reduced training requirements. Such a system could allow even novice laparoscopic surgeons to target relatively small areas of interest in procedures such as biopsies or tumor ablation with relatively high success rates. Additionally, the ability of users to reach targets with no damage to close lying critical structures in more than 50% of cases demonstrates that that the approach could improve safety of laparoscopic procedures and may allow more complex clinical cases to be performed laparoscopically in the future. Increased path lengths and position variation during the targeting of landmarks in the early stages of testing additionally suggests that damage to vessels may also be attributed to lack of tool control and hand eye coordination resulting from the very limited training provided to the subjects. Both of these aspects were, however, seen to improve rapidly over time with very little experience.

The presented targeting evaluation is an initial step towards a quantitative verification of the advantages of endoscopic overlay in laparoscopic surgery. Whilst initial evaluations of the proposed system are promising, further clinical studies, and comparative studies investigating performance compared with standard endoscopic views with both trained and untrained subjects is required before benefits of the approach can be confirmed.

References

[1] Veldkamp, R., Kuhry, E., Hop, W., Jeekel, J., Kazemier, G., Bonjer, H.J., Haglind, E., Påhlman, L., Cuesta, M., Msika, S., Morino, M., Lacy, A.: Laparoscopic surgery versus open surgery for colon cancer: short-term outcomes of a randomised trial. The Lancet Oncology 6(7), 477–484 (2005)

[2] Buell, J., Cherqui, D., Geller, D., O'Rourke, N., Iannitti, D., Dagher, I., et al.: The International Position on Laparoscopic Liver Surgery. Annals of Surgery 250(5), 825–830 (2009)

[3] Marescaux, J., Rubino, F., Arenas, M., Mutter, D., Soler, L.: Augmented-Reality-Assisted Laparoscopic Adrenalectomy. JAMA 292(18), 2214–2215 (2004)

[4] Nicolau, S., Pennec, X., Soler, L., Buy, X., Gangi, A., Ayache, N., Marescaux, J.: An augmented reality system for liver thermal ablation: design and evaluation on clinical cases. Medical Image Analysis 13(3), 494–506 (2009)

[5] Soler, L., Nicolau, S., Hostettler, A., Fasquel, J., Agnus, V., Charnoz, A.: Computer Assisted Digestive Surgery, Computational Surgery and Dual Training, 2010th edn., pp. 139–153. Springer, US (2010)

[6] Hong, J., Sungmin, K., Hashizume, M., Cho, H.: Endoscopic image overlay surgical navigation using augmented and virtual reality technologies. In: Proceedings of 2011 8th Asian Control Conference (ASCC), May 15, pp. 574–578 (2011)

[7] Winne, C., Khan, M., Stopp, F., Jank, E., Keeve, E.: Overlay visualization in endoscopic ENT surgery. International Journal of Computer Assisted Radiology and Surgery 6(3), 401–406 (2010)

[8] Nicolau, S., Soler, L., Mutter, D., Marescaux, J.: Augmented reality in laparoscopic surgical oncology. Surgical Oncology 20(3), 189–201 (2011)

[9] Peterhans, M., vom Berg, A., Dagon, B., Inderbitzin, D., Baur, C., Candinas, D., Weber: A navigation system for open liver surgery: design, workflow and first clinical applications. The International Journal of Medical Robotics + Computer Assisted Surgery: MRCAS 7(1), 7–16 (2011)

[10] Schenk, A., Zidowitz, S., Bourquain, H., Hindennach, M., Hansen, C., Hahn, H., Peitgen, H.O.: Clinical relevance of model based computer-assisted diagnosis and therapy. In: Proceedings of SPIE, vol. 6915(6), pp. 691502\1–691502\19 (June 2008)

[11] Gavaghan, K.A., Peterhans, M., Oliveira-Santos, T., Weber, S.: A portable image overlay projection device for computer-aided open liver surgery. IEEE Transactions on Biomedical Engineering 58(6), 1855–1864 (2011)

[12] Zhang, Z.: A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(11), 1330–1334 (2000)

[13] Li, S., Zhang, G., Wei, Z.: The determination of the intrinsic and extrinsic parameters of virtual camera based on OpenGL. In: Proceedings of SPIE, vol. 6358, pp. 63581C–63581C-6 (2006)

[14] Gavaghan, K.A., Anderegg, S., Peterhans, M., Oliveira-Santos, T., Weber, S.: Augmented Reality Image Overlay Projection for Image Guided Open Liver Ablation of Metastatic Liver Cancer. In: Linte, C.A., Moore, J.T., Chen, E.C.S., Holmes III, D.R. (eds.) AE-CAI 2011. LNCS, vol. 7264, pp. 36–46. Springer, Heidelberg (2012)

[15] Glossop, N.D.: Advantages of optical compared with electromagnetic tracking. The Journal of Bone and Joint Surgery 91(suppl. 1), 23–28 (2009)

[16] Yaniv, Z., Wilson, E., Lindisch, D., Cleary, K.: Electromagnetic tracking in the clinical environment. Medical Physics 36(3), 876 (2009)

[17] Heizmann, O., Zidowitz, S., Bourquain, H., Potthast, S., Peitgen, H.O., Oertli, D., Ket-telhack, C.: Assessment of intraoperative liver deformation during hepatic resection: prospective clinical study. World Journal of Surgery 34(8), 1887–1893 (2010)

[18] Ribes, D., Peterhans, M., Wallach, D., Banz, V., Stavrou, G., Weber, S.: US/CT registration framework for higher instrument guidance accuracy in open liver surgery. In: CURAC 2012, Düsseldorf (2012)

[19] Hansen, C., Wieferich, J., Ritter, F., Rieder, C., Peitgen, H.-O.: Illustrative visualization of 3D planning models for augmented reality in liver surgery. International Journal of Computer Assisted Radiology and Surgery 5(2), 133–141 (2010)

Modeling of Radiofrequency Ablation Lesions for Image-Guided Arrhythmia Therapy: A Preliminary *ex vivo* Demonstration

Cristian A. Linte, Jon J. Camp, David R. Holmes III,
Maryam E. Rettmann, and Richard A. Robb

Biomedical Imaging Resource, Mayo Clinic,
Rochester MN USA
{linte.cristian,robb.richard}@mayo.edu

Abstract. In spite of significant efforts to enhance guidance for catheter navigation, very little has been done to consider the changes that occur in the tissue during ablation as a means to provide feedback on therapy delivery. We propose a technique to visualize the lesion progression and monitor the RF energy delivery by means of a real-time thermal ablation model. The model is based on physical and physiological tissue parameters, and uses heat transfer principles to estimate temperature distribution and geometry of the generated lesion in real time. Validation of the model against experimental measurements recorded in *ex vivo* muscle samples ablated under clinically relevant conditions demonstrated good agreement between the predicted and measured parameters. We believe this technique will enable the generation of real-time thermal maps that can be used to guide the placement of successive lesions to ensure continuous and effective suppression of the arrhythmic pathway.

1   Introduction

Catheter-based ablation therapy has received extensive attention for the treatment of arrhythmic conditions, in particular atrial fibrillation (AF). While the exact etiology of the disease is not entirely known, it is believed that the ectopic foci often originate in the pulmonary veins [1]. However, reported ventures into the treatment of AF have been somewhat discouraging. Studies have shown that 15-40% of patients require revisions [2]. Moreover, it comes with no surprise that the 40-70% success rates for treatment of AF have been accompanied by lengthy procedures, extensive exposure to ionizing radiation, and high risk [3].

Visualization in traditional ablation procedures has been impeded by the inherent limitations of 2D fluoroscopy. The catheter and cardiac silhouettes are visible, but visualization of the endocardial surface and requisite catheter contact is impossible. Extensive research has focused on developing efficient, rapid, robust and intuitive real-time solutions for intra-operative catheter navigation [4,5]. However, most approaches lack the direct spatial and temporal relationship between the available multi-modality data sources and the patient, leaving the clinician with the major task of mentally integrating the multi-modality data,

while manipulating the catheter. Other efforts have leveraged the use of complex imaging technology to enable visualization and monitoring of therapy delivery in real time [6]. Despite the high quality anatomical and functional information, modalities such as CT and MRI are neither efficient nor practical in the interventional suite due to high cost, lack of accessibility, high radiation dose or high magnetic fields, and presence of bulky equipment in the operating room.

Our approach relies on a more practical alternative that allows us to dynamically update, animate, synchronize and visualize the multi-modality data in direct spatial and temporal relationship with the intra-operative patient. To better illustrate the significant value provided by the proposed therapy guidance platform, we follow a hypothetical patient through the workflow associated with this procedure. A dynamic pre-operative CT or MRI dataset is acquired and post-processed to generate a patient-specific left atrial model, which serves as anatomical substrate onto which intra-operative electro-physiology data is superimposed, enabling the identification of the arrhythmic foci for interventional planning [7]. This data is then registered to the patient using previously explored techniques, and further augmented with tracked intracardiac echocardiography (ICE) images, which enables precise catheter positioning for therapy delivery [8] (Fig. 1). Once on target, RF energy is delivered to the tissue under real-time visualization provided to the clinician via a thermal ablation model discussed in this paper .
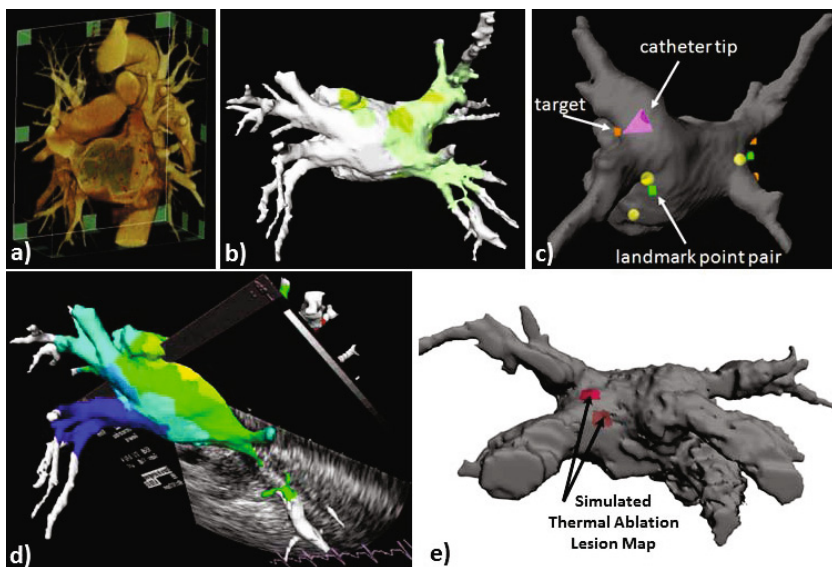
Fig. 1. Advanced platform for image guidance and visualization for left atrial ablation therapy. a) Volume-rendered pre-operative patient-specific CT scan; b) electro-anatomical model showing superimposed activation data, integrated with real-time magnetic catheter tracking (c) and real-time tracked US imaging (d); e) proposed feature addressed in this work for effective representation and visualization of thermal model maps for therapy monitoring and guidance of successive lesions.

This work complements our previous efforts on enhancing catheter guidance by providing a means for estimating and visualizing the effects of RF energy delivery on the tissue. Here we present the development, implementation and initial validation of a thermal ablation model that uses the physiological properties of the tissue and relies on heat transfer principles to predict temperature distribution and lesion progression.

2   Materials and Methods

2.1   Thermal Ablation Model

The response of the tissue to the delivery of RF energy can be approximated with sufficient accuracy by a coupled resistive - conductive heat transfer process. The proposed thermal ablation model will incorporate a resistive component occurring at the catheter tip-tissue interface, coupled with a purely conductive component responsible for the diffusion of thermal energy into the tissue.

Governing Principles. The proposed thermal model follows a finite element formulation of the modified Pennes bioheat equation [9], implemented and solved using a 3D image-based approach,

$$\rho_t \, c_t \, \frac{dT}{dt} \;=\; \nabla \cdot (k \, \nabla T) \;+\; \sigma \, |\nabla V|^2 \;-\; \rho_b \, c_b \, \alpha(t) \, \omega \, (T - T_{amb}) \;+\; Q_m \,, \quad (1)$$

where ρ_t is the tissue density (1060 kg/m^3) [10]; c_t is the tissue specific heat (3600 J/kg·K) [10]; σ is the tissue's temperature-dependent electrical conductivity (S/m); V is the electric potential (V); k is the initial tissue thermal conductivity (0.42 W/m·K) [10]; ρ_b is the blood density (1000 kg/m^3) [11]; c_b is the blood specific heat (4180 J/kg·K); α is a tissue state coefficient (ranging between 0 and 1, depending on the extent of tissue damage); ω is the blood perfusion coefficient [11], which for the case of left atrial applications, it can be ignored due to negligible perfusion from the vasculature in the region; T_{amb} is the ambient body temperature (37° C); and Q_m is the metabolic heat source term, which is insignificant [12]. The spatial resolution of the computational model is dictated by the spatial resolution of the image based on which the image-based formulation is implemented. For the *ex vivo* study described below, the resolution of the computational grid is 0.5 mm isotropic. Moreover, the computational time step is 1 second.

The time-varying density, thermal and electrical conductivity of tissue are considered in the model, along with their changes due to dehydration (i.e. with dehydration, tissue becomes more insulative, therefore experiencing reduced thermal conductivity). Internal tissue temperature is estimated by updating the thermal and electrical conductivity and the position of the ablation catheter configuration on a voxel-by-voxel basis for the duration of the ablation. The ablation catheter can be represented by either an image-derived object segmented from a

3D US image, or by a virtual construct based on its properties; various catheter shapes and poses can be evaluated against patient-specific tissue structures. The cooling effect of the endocardial blood flow is also accounted for via the perfusion coefficient $\alpha(t)$.

Computational Implementation. The model is implemented based on a pre-acquired anatomical image volume which is first thresholded and labeled as blood and tissue regions, and initialized by assigning electrical and thermal conductivity values to each region at body temperature (37°C). During clinical ablation, the RF generator is typically operated in the temperature-controlled mode: delivered RF energy is increased up to the set maximum power level until the preset target electrode temperature is reached. Power is then scaled back to maintain this temperature to prevent tissue charring.
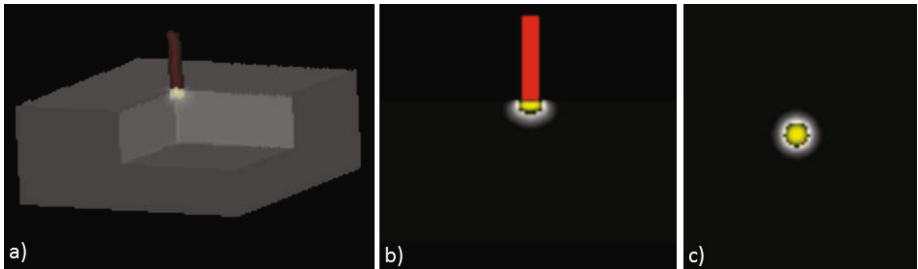
Fig. 2. Ablation procedure portrayed using the image-based model: a) ablated tissue sample and created lesion (in yellow) and penumbra (white) shown from two viewpoints (b & c)

The ablation catheter is currently represented by a virtual construct based on its physical properties. Its interaction with the tissue is defined via a voxel-based occupancy within the image volume. During each computational cycle, the catheter-representing voxels are "heated" to the target electrode temperature under maximum power level heat-flux, which is then damped to maintain the temperature for the duration of the delivery. The voxels adjacent to the catheter undergo resistive heating, while the voxels sufficiently remote from the catheter undergo an anisotropic conductive heat transfer process. After each computational cycle, the electrical and thermal conductivity of the proximal and distal voxel regions, respectively, are dynamically updated to reflect the physiological tissue changes caused by dehydration, and the temperature distribution within the tissue is dynamically estimated. Prior to proceeding to the next computational cycle, a fresh supply of blood flowing through the left atrium is modeled by re-initializing the blood pool boundary condition to 37°C, followed by a reheat of the voxel-based catheter-representation to the target electrode temperature.

Predicted Temperature and Lesion Size Measurements. This approach allows the formulation of model-predicted lesions built based on the cumulative "heating" effect on the voxels that reach temperatures above the physiologically-established threshold cell-death temperature. The cumulative tissue damage is computed after each computational "heating" cycle and expressed as the total number of voxels that have been exposed to temperatures above the cell-death threshold for a certain duration, typically 5s. The effect of tissue damage is then incorporated into the model via the tissue state coefficient. Exposure is modeled in a reversible fashion, allowing incompletely ablated voxels to rehydrate if not exposed to temperatures above the cell-death threshold for sufficient time, mimicking temporarily stunned tissue (i.e., lesion penumbra). Following a certain number of "heating" cycles, voxels undergoing sufficient exposure are digitally marked to represent the created lesion.
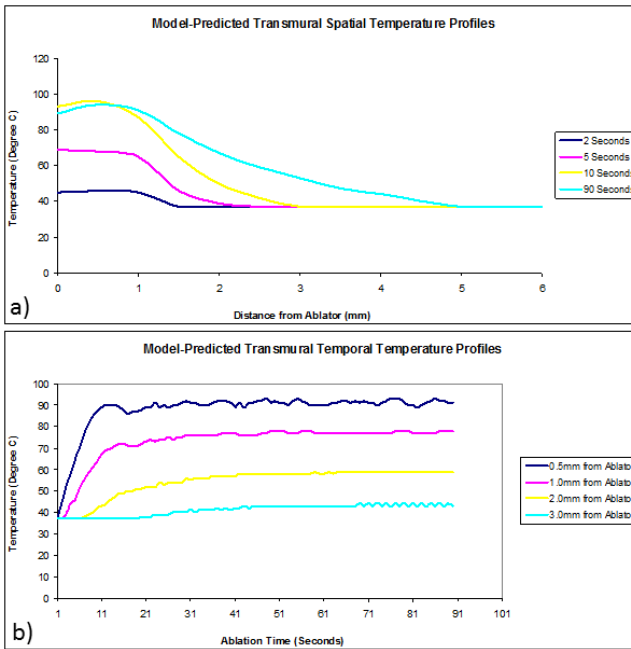
Fig. 3. a) Model-predicted transmural temperature profiles at several stages during ablation; b) Time-varying model-predicted temperature profiles at different distances from the electrode tip

2.2   Experimental Design

In addition to modeling the thermal ablation process, we have also conducted a series of experiments using lean *ex vivo* beef muscle samples. The samples were prepared and submerged in a convective 0.9% saline bath maintained at a constant temperature of 37°C. An EPT-1000 XP (Boston Scientific, Natick, MA)

RF power generator was used to deliver ablation burns using a 7F electrode catheter with temperature feedback. The generator was operated in the temperature controlled mode, with a 70°-90°C electrode target temperature and 40W maximum power level. To mimic clinical practice, the ablation exposure time was set to 60 s and typical impedance values averaged 80-100Ω. To characterize tissue temperature during ablation for validation purposes described later, focal fiberoptic temperature probes (OpSens Inc., Quebec City, Canada) were inserted within the tissue at prescribed distances and depths relative to the ablation electrode (Fig. 3). Each sensor provided a precision on the order of 0.1 C with a 50 Hz sampling rate.
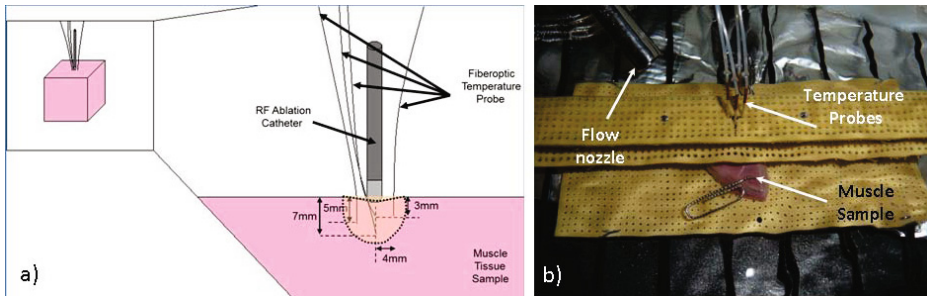
Fig. 4. a) Schematic diagram of the experimental setup and temperature measurements during the RF ablation of *ex vivo* muscle samples; b) Experimental setup showing temperature probes and electrode inserted in muscle submerged in a convective saline bath at 37°C to simulate physiological conditions. A grid template is used to ensure probe insertion at desired locations.

3 Experimental Validation, Evaluation and Results

We now show the results of two sets of ablation experiments. The target electrode temperature was set to 90°C and 60°C. Using a grid template of known dimensions, four fiberoptic temperature probes were inserted into the tissue sample at 2.5 mm and 5 mm away (R) from the electrode and 3 mm depth (D). The probes were inserted in a cross pattern — two probes at 2.5 mm R x 3 mm D, on opposite sides of the electrode, and the other two at 5 mm R x 3 mm D. Temperature measurements were simultaneously recorded from all probes during the 60 second ablation procedure. Similar parameters were used to configure the thermal model, and temperature measurements at the same spatial locations were obtained.

Table 1 provides a summary of the experimentally recorded and model-predicted data (six time-points during the ablation, two spatial locations and two target electrode temperature conditions) for the 90°C ablation procedure. As noted, the results show a less than 3°C difference between the predicted and measured temperature consistently for the entire duration of the ablation.

Table 1. Model predicted vs. experimentally measured temperatures at 2.5 mm and 5.0 mm radius (R) and 3 mm depth (D) for the temperature-controlled ablation procedure at 90°C. Note: Measured values were averaged between two symmetrically places sensors. Predicted values were recorded at nearest second to measurement time.

Time	2.5 mm R x 3.0 mm D		5.0 mm R x 3.0 mm D	
(sec)	Model T (°C)	Exprmtl. T (°C)	Model T (°C)	Exprmtl. T (°C)
11.2	64	61.6	37	37.1
19.6	72	70.7	39	38.8
30.8	77	74.9	41	39.7
39.2	78	77.2	43	40.4
50.4	79	78.5	43	40.8
58.8	80	78.9	44	41.4

The agreement between the model and experimental data can be observed in the time-varying temperature plots displayed in Fig. 4e, along with the electrode temperature plot.

Fig. 4 (90°C target electrode temperature) shows the predicted lesion, as well as the temperature measurements at the end of the ablation procedure at both spatial locations. While the proximal probe (Fig. 4a) locations (2.5 mm R x 3 mm D) achieved temperatures above cell-death threshold (∼55°C), the distal (Fig. 4b) locations (5 mm R x 3 mm D) did not. These results are confirmed by the observed physical lesion (Fig. 4c&d) size (∼ 5.5 mm diameter), which places the first recorded location in the lesion, and the second in the lesion penumbra.

Table 2. Model predicted vs. experimentally measured temperatures at 2.5 mm and 5.0 mm radius (R) and 3 mm depth (D) for the temperature-controlled ablation procedure at 60°C. Note: Measured values were averaged between two symmetrically places sensors. Predicted values were recorded at nearest second to measurement time.

Time	2.5 mm R x 3.0 mm D		5.0 mm R x 3.0 mm D	
(sec)	Model T (°C)	Exprmtl. T (°C)	Model T (°C)	Exprmtl. T (°C)
11.2	44	43.2	37	36.7
19.6	48	49.8	37	37.5
30.8	51	52.7	37	38.2
39.2	53	54.3	37	39.1
50.4	57	55.3	37	39.0
58.8	61	56.3	37	39.1

Table 2 provides a summary of the experimentally recorded and model-predicted data (six time-points during the ablation, two spatial locations and two target electrode temperature conditions) for the 60°C ablation protocol.

Fig. 5. Target Electrode Temperature: 90°C. a & b) Model predicted lesion showing the spatially samples locations corresponding to the inserted probe locations; c & d) Physical ablation lesion ($\sim$ 5.5 mm diameter); e) Time-varying temperature plot of the electrode, as well as model-predicted and measured temperatures at proscribed spatial locations.

The agreement between the model and experimental data can be observed in the time-varying temperature plots displayed in Fig. 5e, along with the electrode temperature plot.

Fig. 5 (60°C target electrode temperature), on the other hand, shows that the proximal probes location achieved temperatures in the low 40°C, while the temperature at distal probe locations did not rise above the ambient temperature (hence the red background at 37°C). The captured images of the physical lesion also show a much lesser extend of tissue damage, with no actual lesion created (temperatures below the cell-death threshold), but rather resulting in a stunned tissue region, often referred to as edema in the clinical literature [13].
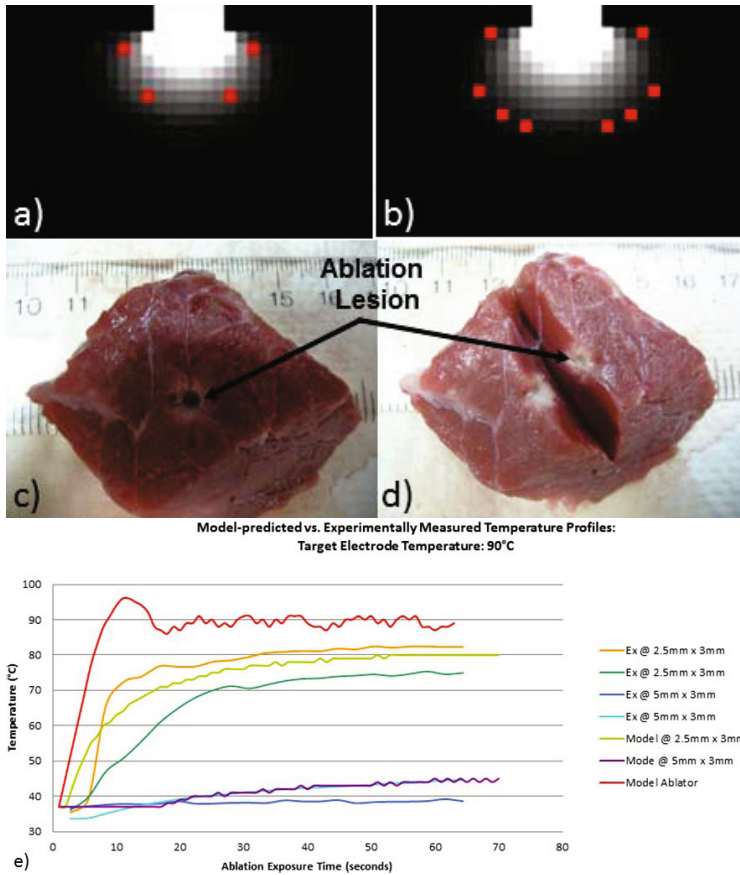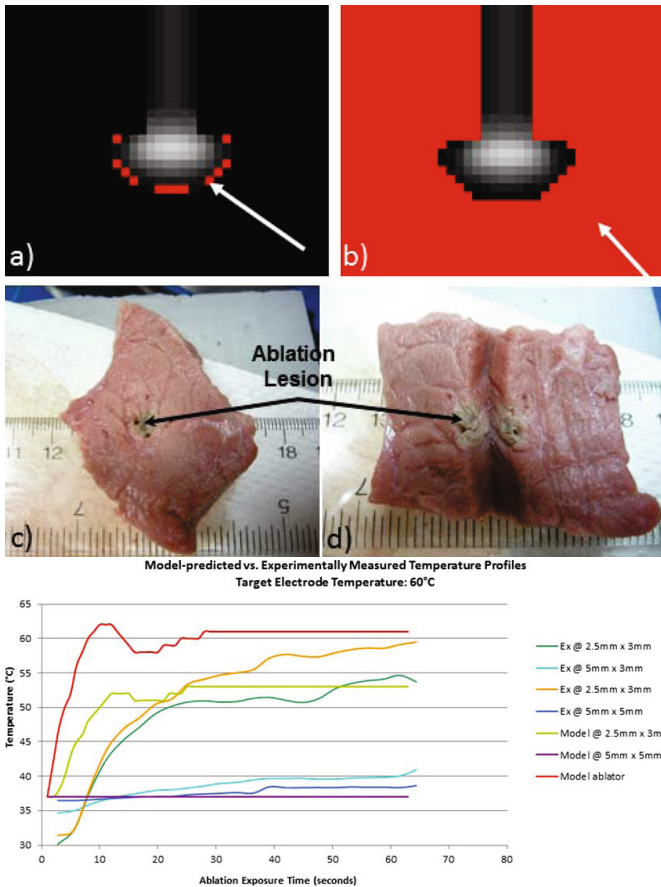
Fig. 6. Target Electrode Temperature: 60°C. a & b) Model predicted lesion showing the spatially samples locations corresponding to the inserted probe locations; c & d) Underdeveloped physical ablation "lesion", showing much lesser extent of tissue damage; e) Time-varying temperature plot of the electrode, as well as model-predicted and measured temperatures at proscribed spatial locations.

4   Discussion

This work describes the initial development of a 3D real-time image-based thermal model that enables monitoring and visualization of the tissue response during ablation exposure, including achieved temperature distribution, formation and progression of the lesion. The studies conducted in *ex vivo* beef muscle have demonstrated less than 5°C differences between the model-predicted and experimentally measured temperature profiles. The predicted and observed lesion patterns were also in agreement, revealing both the core scar region characterized by permanent tissue damage, as well as the surrounding penumbra of the ablation scar, suggestive of temporary stunning of the tissue with the potential

of function restoration. These lesions could also be visualized via real-time US images already available for intra-operative guidance, as scarred tissue expresses a different appearance in the US images than viable tissue [14].

We have also identified several limitations of this study with respect to a ream cardiac ablation procedure. One limitation consists of the constant electrode-tissue contact achieved *ex vivo*, which would be difficult to achieve *in vivo* due to the cardiac motion. The inherent variability of the tissue parameters, which is difficult to assess, would also lead to uncertainties and inherent variations in the predicted temperature and lesion progression; however, exact knowledge of the tissue temperature is not needed, but rather sufficient documentation that temperatures above cell-death temperature (50° C) have been achieved. Lastly, the *ex vivo* study described here has been performed using a non-irrigated catheter, approach which can be considered somewhat outdated, as research has shown that irrigated catheters enable delivery of higher RF power and hence better lesion development. However, we plan to adapt the current model to replicate the results achieved using a irrigated-tip catheter; this condition can be implemented by limiting the target temperature that the electrode tip is allowed to reach before power delivery is reduced.

As far as the visualization and display of the model-predicted temperature distribution and ablation lesions, 2D slice renderings of volumetric data are commonly used in the clinical routine. In addition, several other approaches have been explored by other groups [15], including volumetric overlays, multi-parameter contours showing the representative isothermal lines (typically the 50° C isotherm is a commonly used approach), as well as 3D volumetric visualizations using combined volume rendering of the lesion and substrate tissue.

As part of our future work, a series of animal studies will be conducted *in vivo*, where the delivered lesions characterized post-procedure via DCE-MRI can be assessed against those predicted by the thermal ablation model. Lastly, this approach will allow for the identification and characterization of endemic scar (i.e. tissue swelling rather than a fully developed scar) and correlated back to the model-predicted lesions.

Knowledge of the true sites of the ablation lesions and their characterization as necrosis or edema is of significant importance and will help validate the proposed thermal ablation model via a series of *in vivo* animal studies. The extent of necrosis as a measure of irreversible damage has been shown to be in good agreement with delayed contrast-enhanced (DCE) MR images for RF ablation lesions [13]. Edema, on the other hand, is an inflammatory response to injury known to occur after ablation, which subsides after a matter of weeks, with concomitant restoration of electrical connection. Several studies have reported that edema also overlaps with the areas of irreversible injury [13].

With the aid of the presented thermal model, the cardiologist will visualize in real-time the changes induced in the tissue via RF energy delivery, adjacent to the catheter guidance information. The tracked catheter tip dictates the location where the ablation lesion should be displayed. The physiological changes occurring in the tissue are displayed by means of temperature colour maps "painted"
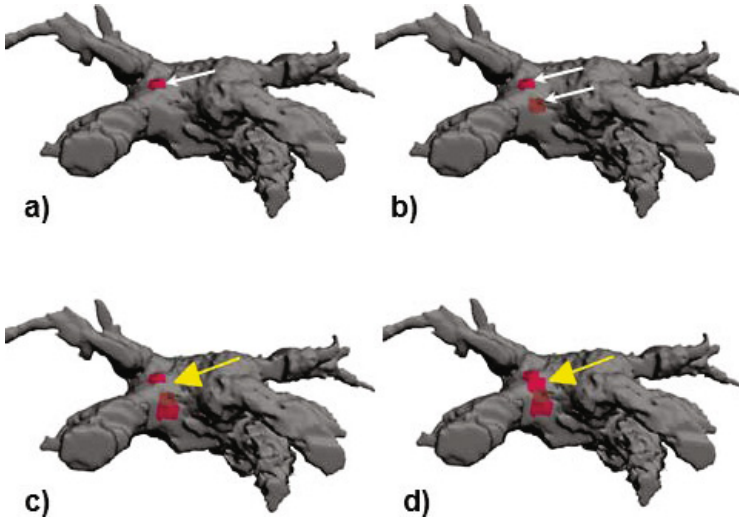
Fig. 7. An illustration of the progression of target ablations around a pulmonary vein using the thermal map model. a) A first ablation lesion has been created; b) A second lesion has been created under shorter exposure; c) A third lesion is created, however too far from the previous (gap shown by arrow); d) The gap filled by creating another lesion under guidance provided by the thermal map model, ensuring a continuous burn pattern.

onto the pre-operative image. The dynamically updated lesion maps will indicate where subsequent burns should be placed, as well as whether additional exposure is required to achieve desired lesion size, ensuring effective electrical isolation (Fig. 6).

5   Conclusion

In summary, this work describes the initial development and validation of an endocardial tissue thermal ablation model. Unlike other commercially available packages which focus on modeling the tissue in its full complexity at a high computational expense, our proposed solution is based on an image-based implementation and provides real-time visualization of temperature distribution and lesion progression.

References

1. Blaauw, Y., Crijns, H.J.: Atrial fibrillation: insights from clinical trials and novel treatment options. J. Intern. Med. 262, 593–614 (2007)
2. Pappone, C., et al.: Circumferential radiofrequency ablation of pulmonary vein ostia: A new anatomic approach for curing atrial fibrillation. Circulation 102, 2619–2628 (2000)

3. Oral, H., Pappone, C., et al.: Circumferential pulmonary vein ablation for chronic atrial fibrillation. N. Engl. J. Med. 354, 934–941 (2006)
4. Sra, J., Krum, D., et al.: Registration of 3D left atrial CT images with fluoroscopy. Heart Rhythm 2, 1020–1022 (2000)
5. Knecht, S., Skali, H., et al.: CT-fluoroscopy overlay evaluation during catheter ablation of left atrial arrhythmia. Europace 10, 931–938 (2008)
6. Dickfeld, T., Calkins, H., et al.: Stereotactic catheter navigation using magnetic resonance image integration in the human heart. Heart Rhythm 2, 413–415 (2005)
7. Rettmann, M.E., Holmes III, D.R., Cameron, B.M., Robb, R.A.: An event-driven distributed processing architecture for image-guided cardiac ablation therapy. Comput. Methods Programs Biomed. 95, 95–104 (2009)
8. Linte, C.A., Lang, P., Rettmann, M.E., Cho, D.S., Holmes III, D.R., Robb, R.A., Peters, T.M.: Accuracy considerations in image-guided cardiac interventions: experience and lessons learned. Int. J. Comput. Assist. Radiol. Surg. 7, 13–25 (2011)
9. Pennes, H.H.: Analysis of tissue and arterial blood temperatures in the resting human forearm. J. Appl. Physiol. 85, 5–34 (1998)
10. Duck, F.: Physical Properties of Tissues: A Comprehensive Reference Book, pp. 167–223. Academic Press, New York (1990)
11. Tungjitkusolmun, S., Staelin, S.T., et al.: 3D finite element analyses for RF hepatic tumor ablation. IEEE Trans. Biomed. Eng. 49, 3–9 (2002)
12. Chang, I.A., Nguyen, U.D.: Thermal modeling of lesion growth with RF ablation devices. Biomed. Eng. Online 3, 1–19 (2004)
13. Knowles, B.R., Caulfield, D., et al.: 3-D visualization of acute RF ablation lesions using mri for the simultaneous determination of the patterns of necrosis and edema. IEEE Trans. Biomed Eng. 57, 1467–1475 (2010)
14. Chan, R., Packer, D.: Accuracy of ICE measurement of RF ablation lesion dimensions in the intact canine atrium and ventricle. Circulation 92, I–794 (1995)
15. Rieder, C., Kroger, T., Schumann, C., Hahn, H.K.: GPU-based real-time approximation of the ablation zone for radiofrequency ablation. IEEE Trans. Vis Comput. Graph. 17, 1812–1821 (2011)

Visualization of Organ Motion during Breathing from 4D Datasets

Markus Müller, Athanasios Karamalis, and Nassir Navab

Computer Aided Medical Procedures
Technische Universität München, Germany

Abstract. The introduction of $3D + t$ medical datasets is posing new challenges to visualization. Volumetric and flow visualization are established fields offering a wide spectrum of techniques for visualizing $3D + t$ datasets. In this work we address the problem of visualizing the motion of organs during breathing. As opposed to flow visualization we are not directly interested in the underlying flow, but in the deformation caused by the flow. Therefore, visualization of breathing motion focuses on emphasizing the organ motion while preserving the anatomical context provided by the volumetric visualization. In this work we will discuss methods from flow and volume visualization, their applications, and introduce alternative visualization approaches for enhancing the perception of organ motion due to breathing.

Keywords: 4D Visualization, Volume Rendering, Flow Visualization, Deformation, Breathing.

1   Introduction

Advancements in medical imaging do not only improve the quality of the images, but also increase the amount of data available to the physicians, a trend that is likely to continue in the future. Visualization has always played a crucial role in presenting the available data in a coherent and intuitive way, assisting physicians and clinical personnel in diagnosis, planning, and treatment. The role of visualization becomes even more important with higher-dimensional datasets like for example high-angular resolution diffusion imaging (HARDI). These complex datasets can only be interpreted with the help of sophisticated visualization techniques [15].

Relatively recent advancements in medical image acquisition allow us to acquire multiple volumetric datasets over time $(3D + t)$, which we will also refer to as 4D datasets. These datasets usually come from 4D ultrasound, 4D computed tomography (CT), and 4D magnetic resonance imaging (MRI). The established fields of volumetric and flow visualization offer a variety of techniques for visualization of $3D + t$ datasets. Visualization of volumetric datasets has constantly improved in the previous years as a result of hardware and algorithmic advancements [3]. The field of flow visualization has a long history and introduced sophisticated methods for the visualization of high-dimensional datasets. For a

more comprehensive review and categorization on flow visualization we refer to Laramee et al. [13].

The visualization of organ and tissue motion is introducing new challenges. Although, flow visualization techniques are directly applicable for the visualization of flows in volumetric medical datasets they do demonstrate limited applicability in visualizing organ and breathing specific motion. In this work we are not directly interested in visualizing the flow in $3D + t$ medical image datasets, but in visualizing the effect of the flow, i.e., the tissue and organ motion subject to the flow. Medical education and training is one of the major application areas of such visualizations, which makes it important to preserve the anatomical context in the visualization. In the next sections we will discuss the applicability of methods from volume and flow visualization to our problem. Furthermore, we will introduce an alternative visualization approach that emphasizes motion through the volume rendering transfer function.

2   Previous Work

In flow visualization we are mostly concerned with visualizing steady and unsteady flows. Laramee et al. [13] categorized flow visualization into direct, texture/dense, and geometry visualization. Direct visualization is performed with the help of glyphs, most commonly arrows, that follow the direction of the flow and can encode further properties like magnitude. However, these visualizations become rapidly overwhelming with increasing number of glyphs required for the visualization of detailed and complex flows. Texture/dense approaches provide flow information for every pixel. The most popular approach is the line-integral-convolution (LIC) method [13], which convolves a noise texture with the flow. The disadvantage of this method is that it is difficult to distinguish specific characteristics of the flow. The geometry approaches make use of stream, streak, and path lines [14]; and offer a compromise between direct and dense flow visualization by emphasizing important characteristics of a flow.

Ray-casting has been established as the standard for volumetric visualization, especially with the introduction of GPU accelerated ray-casting [12]. Transfer functions are playing the most important role for providing the desired visualization as they allow the implicit segmentation of anatomical structures of interest interactively during rendering. Volume rendering has the benefit of providing visualization directly derived from the volumetric data set; preserving the anatomical context with the use of appropriate transfer functions. For a more comprehensive overview we refer to Engel et al. [3].

The visualization of $3D + t$ medical datasets has already been addressed for different application domains, however, to the best of our knowledge no work is specifically addressing the problem of visualizing organ and breathing specific motion. The closest work we found was from Handels et al. [7] focusing on computing tumor mobility for radiation therapy. The motion of the lung and the tumor were visualized with vectors and color coded meshes. A formal discussion on visualization of organ motion due to breathing was not presented. We believe
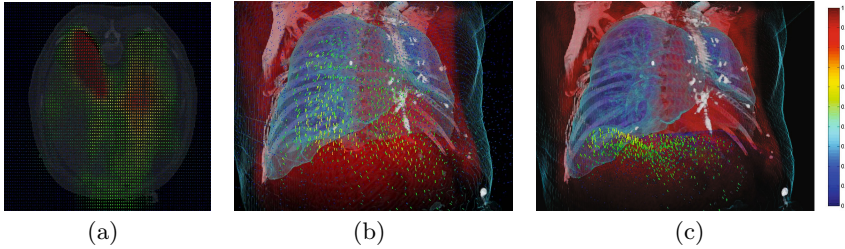
Fig. 1. Direct visualization of motion field with vectors. The vector colors encode the normalized motion magnitude. In (a) a MPR slice of a volume is displayed together with the motion vectors. In (b) the vectors are combined with the volume rendering of the dataset (from now on referred to as *vector* visualization). In (c) only the vector spatially belonging to the liver are displayed.

that the limited number of publications dealing with this subject is related to the relatively recent public availability of $3D + t$ datasets to the visualization community. Fortunately, projects like the POPI model from the Léon Bérard Cancer Center & CREATIS lab (Lyon, France) [17] provide not only the 4D CT datasets, but also the displacement fields with respect to a reference volume, which are computed by parametric and non-parametric deformable registration methods.

Visualizing organ motion on a cutting plane of the volumetric dataset can be adequately achieved with standard flow visualization techniques [14,13]. Figure 1(a) shows such a rendering. It is evident that rendering only a 2D cutting plane results in losing the anatomical context in the visualization. Furthermore, out-of-plane motion can not be visualized with this approach. For that reason, this work will only discuss 3D visualizations using volume rendering as the technique of choice for providing fast and anatomically preserving volumetric visualizations [3,12]. The new challenge is how to incorporate the visualization of organ specific motion into volume rendering. In the next sections we will discuss the applicability of well-known visualization techniques for visualizing the effect of breathing motion and introduce a new technique for emphasizing the motion magnitude through the transfer function.

2.1   Direct Volume Rendering (DVR)

Ray-casting has been established as the standard for volumetric visualization, especially with the introduction of GPU accelerated ray-casting [12]. For a more comprehensive overview we refer to Engel et al. [3]. The 4D CT POPI dataset - used in this work - is essentially a collection of 3D CT volumes acquired over different breathing stages. Sequential volume rendering of these already allows a decent visualization of the motion. A frame from this visualization is shown in Fig. 2(a) and 3(a).

2.2   DVR and Vector Geometry

The availability of a displacement field allows the integration of direct flow visualization techniques. These dense displacement fields can be computed with deformable/non-rigid registration techniques. There the goal is to compute a dense deformation field that maps voxels/pixels from one dataset to the corresponding ones in another dataset acquired at a different time. The field of image registration has been studied for almost 30 years and numerous methods have emerged for deformable registration [5]. In this work we use the pre-computed displacement fields available for each volume of the POPI dataset [17], which were computed with a free-form deformation approach [2] and a Demons-based approach [16]. Alternatively, deformable registration could be computed with publicly available software packages, including Elastix [10] and Drop [4].

In the field of flow visualization a vector field is often visualized by the use of arrows; encoding both direction and magnitude of the flow [14]. Combining this visualization with DVR allows the examination of the exact deformation for different anatomical structures while preserving the context. A realistic occlusion handling of the vectors does greatly enhance depth perception of the geometry within the volume rendering. Occlusion handling was implemented in our renderer, whereas, efficient techniques for this are further discussed in [3]. A frame from this visualization is shown in Fig. 1(b). When using this direct vector visualization to display the motion of the complete volume with high detail, the final image is cluttered with too much information so that the user is not able to perceive it properly. Increasing the spacing between the vectors would lead to less elements and clutter, but the information between those elements is not rendered. This problem can be partially addressed by rendering only selective regions of interest like in Fig. 1(c), where arrows are only rendered for the liver region, which can be determined by segmentation. Throughout this work the random walks image segmentation method was used [6] for semi-automatic segmentation of organs of interest. In general, the direct flow visualization approaches suffer from occlusion of the geometry and cluttering of the scene, thus, preventing organ specific motion perception.

2.3   DVR and 3D Geometry

The direct vector visualization is useful if the exact movement of a local element is of interest. However, for medical visualization we are also highly interested in how the internal anatomy is affected by the breathing motion and how organs move in relation to each other. Transfer functions in DVR allow an implicit segmentation of the anatomical regions of interest. However, these are often too fuzzy to recognize small deformations. The relative motion of organs to each other and detailed motion on the surface of organs can be highlighted when using abstraction of these objects in form of 3D polygonal meshes. These can be generated prior to visualization from volume segmentations. If a displacement field is available then the segmentation is performed once and is propagated by the field over time. A visualization with a mesh of the liver is shown in Fig. 2(b).
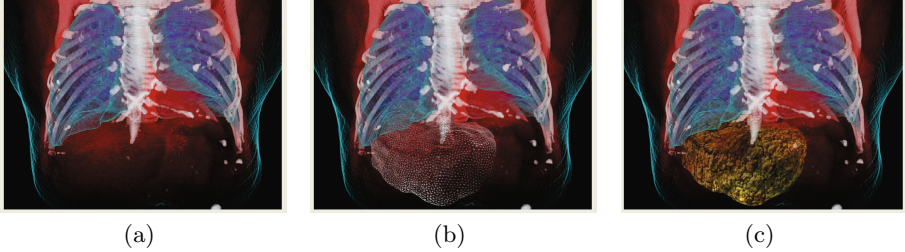
(a)   (b)   (c)

Fig. 2. Volume rendering combined with mesh rendering for emphasizing organ spe-
cific motion. In 2(a) a reference VR is provided (from now on referred to as *direct*
visualization), 2(b) shows a wireframe rendering of the mesh (from now on referred to
as *wireframe* visualization), in 2(c) the mesh is textured with a dense/noisy texture,
mimicking cirrhotic liver.

In order to increase the perception of the mesh/organ distortion we use a noisy
texturing as shown in Fig. 2(c), which has a similar perceptual effects as LIC
techniques.

2.4   Motion Transfer Function

The transfer function (TF) is playing the most important role for providing the
desired visualization. A transfer function maps physical values from the volume
to a more appealing value for visualization, using transparency to separate in-
teresting anatomy and color to emphasize different organs and structures. The
design of a good transfer function is a field of research by itself and often left to
the user through interaction [3].

Here we introduce a new approach of encoding the motion in the transfer
function of the volume rendering, that we denote for the purpose of this paper
motion transfer function (MotionTF). More specifically, the information from the
deformation field is encoded in the TF. Formally, for volume rendering lets define
the emissive color $c(t) := c(s(\mathbf{x}(t)))$ and absorption coefficient $k(t) := k(s(\mathbf{x}(t)))$
where $s(\mathbf{x}(t))$ is the scalar value in the volume at position $\mathbf{x}(t)$, t is the distance
from the ray origin, and $c : X \to (R, G, B)$, $k : X \to A$ are the color and opacity
transfer functions defined by the user. The continuous volume rendering integral
that determines the total amount of radiant color/energy C is given as:

$$C = \int_0^\infty c(t) \cdot \exp^{- \int_0^\infty k(s)ds} dt \tag{1}$$

In the discrete implementation the integral is usually substituted by a Riemann
summation. For a more comprehensive overview on volume rendering theory and
implementation we refer to [3]. In this work we integrate the information from
the deformation field d into the color TF. More specially, we define:

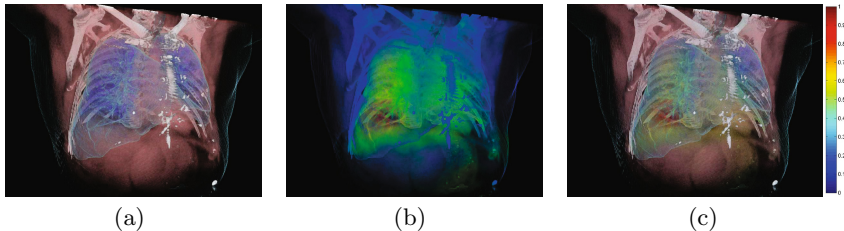$$c_m(t) = m(\|\nabla d(x(t))\|) \tag{2}$$

Fig. 3. In (a) we see a reference volume rendering using a 1D transfer function. In (b) the motion magnitude is encoded directly on the volume with a transfer function (from now on referred to as *MotionTF* visualization). In (c) the reference volume rendering (a) is blended with the motion color volume rendering (b) (from now on referred to as *blended* visualization).

where $m : X \rightarrow (R, G, B)$ is the transfer function mapping the deformation field magnitude to a color. Substituting the original color TF in Eq. 1 with the new motion TF (c_m) results in the visualization shown in Fig. 3(b). A complete sequence of this visualization method is further shown in Fig. 4. The c_m function can be user defined to emphasize motion depending on its magnitude. The presented visualization method is especially well-suited for assessing the global deformation; a perceptual benefit that is also confirmed in our survey later on. Additionally, a blending of the original and motion TF can be used to preserve further the anatomical context provided from the standard color TF, as shown in figure 3(c). This is achieved by substituting the original $c(t)$ with $c'(t) = (1 - \alpha)c(t) + \alpha c_m(t)$, were α is user-defined blending weight. The transfer functions should use distinct colors palettes, preferably with different saturations, in order to avoid the blending of similar colors from the color and motion TF.

3   Survey

We performed a survey with 15 participants that are active researchers in the field of medical image computing, thus, have experience with both medical datasets and visualization. Evaluating the effectiveness of visualization methods remains difficult [9] and we chose to restrict our questions both in number and scope to reduce the bias in the survey [11]. Accordingly, all participants were given videos of five visualization methods on the POPI dataset, which they could view as much as they liked (all images in this work are taken from these videos). Following questions were asked: Which visualization method shows best the: (*Q1*) magnitude of the motion, (*Q2*) direction of the motion, (*Q3*) the interaction/relative motion between anatomies/organs, and (*Q4*) the detailed motion inside the organs. Consequently, the method of choice was recorded for each participant and for each question; summarized in Table 1.

The survey answers for *Q1* indicate that our method (MotionTF) is very well suited for visualizing the magnitude of the motion in the dataset. This was
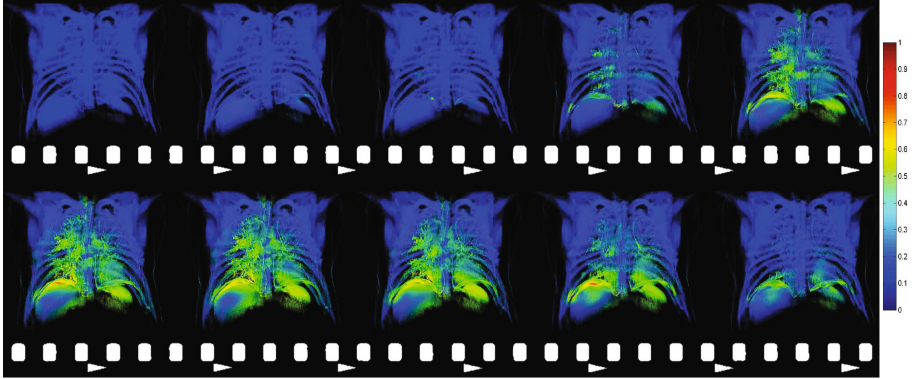
Fig. 4. Images show a sequence from the rendering of the 4D CT dataset at different timesteps/breathing stages, using our motion transfer function

Table 1. Each cell contains the number of times participants identified a method as the best for a specific question (*Q1-Q4*)

	Vector	Direct	Wireframe	MotionTF	Blend
Q1	0	0	1	**14**	0
Q2	5	2	6	2	0
Q3	1	0	**12**	2	0
Q4	2	0	1	7	5

expected given that the method was designed to specifically encoded the motion magnitude. The answers to the questions regarding the direction (*Q2*) and the detail of the motion inside the organs (*Q4*) do not allow clear conclusions to be drawn. Furthermore, the fact that users had different preferences for these tasks shows the difficulties in defining visualization approaches with a clear and common perceptual effect for all users for medical 4D visualization. Last but not least, the wireframe visualization, addressed in *Q3*, proves especially useful for showing organ related motion. The clear delineation with the wireframe geometry proved to be a very important aspect for this result.

4   Discussion

To provide interactive framerates we utilized modern GPU programming APIs including: OpenGL, GLSL, and CUDA. Improving the overall quality of the volume rendering could further improve the perception of motion in the datasets, see [3] for high-quality volume rendering techniques. Furthermore, virtual clipping planes can also assist in defining specific regions of interest in the visualization. The presented MotionTF is a specific case of a visualization for breathing
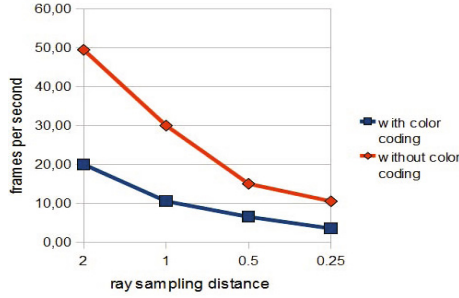
Fig. 5. performance of the MotionTF

motion. Alternative flow information could also be encoded in the function like the direction of the flow, as commonly used in 2D optical flow visualization [1].

The visualizations where tested on a Intel Core 2 Quad CPU with 4 cores running at 2.50Ghz, 4 GB of RAM and a NVIDIA GeForce GTX260 with 896 MB VRAM. Most of the resources are used by our very basic implementation of the volume renderer which is running with around 30 frames per second (fps) and a resolution of 800x600. Compared to the raycasting, the geometry of the direct vector or mesh visualizations with around 100.000 vertices had no notable impact on the performance (around 1-2 fps). The MotionTF on the other hand needs two additional texture lookups per sampling point (one for the vector-field and one for the second transfer function), which approximately reduces the performance of our test system by half depending on the sampling rate (see Fig. 5).

The presented survey focused on the evaluation of basic perceptual questions. We chose videos instead of interactive use of the renderer to ensure the same experience for all users. A clinical survey towards the evaluation of application specific visualizations with the participation of multiple physicians would be of high interest. Possible applications could involve the assessment of the change in organ motion due to illness. Specific examples include the change of elasticity of the lung due to asbestosis, fibrosis, and in general scar tissue. However, a prerequisite for such an evaluation is the availability of a sufficiently large evaluation dataset, which was not available for this work. The visualization techniques presented are generalizable to other modalities including MRI and US. However, separate user-studies are required to analyze the perceptual benefit. Especially US is challenging for visualization because of artifacts and speckle [8], and would require adjustments to transfer function design and blending.

We performed first experiments for visualization of 4D ultrasound acquired from different breathing stages. First results are presented in Fig. 6. Furthermore, a sequence of the 4D visualization, using the introduced MotionTF and wireframe visualization, is shown in Fig. 7. Again, semi-automatic segmentation [6] of the liver anatomy was obtained before visualization to generate the
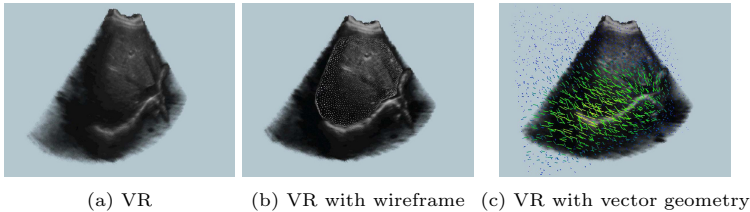
(a) VR   (b) VR with wireframe   (c) VR with vector geometry

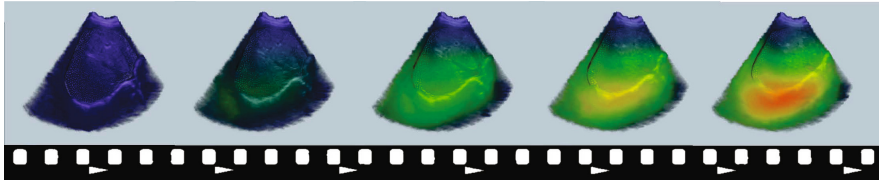**Fig. 6.** Examples of 4D ultrasound visualization

Fig. 7. Example of 4D ultrasound visualization sequence using MotionTF method

required mesh geometry. The acquisition of the dataset was performed with a so-called ultrasound wobble probe; requiring image-based gating to ensure image consistency in acquired volumes [18]. The deformation field was computed with the method introduced in [19]. These initial results demonstrate that the MotionTF provides good perceptual information of motion magnitude in the dataset. Future research on improving transfer functions for ultrasound volume rendering, combined with the introduced visualization methods, could notably improve the overall perceptual benefit.

5   Conclusion

In this work we addressed the specific challenges in the visualization of breathing motion. Furthermore, we presented a new, simple, and effective approach for visualizing the global magnitude of the motion directly in the volume rendering using the transfer function. Flow visualization offers a wide array of methods for visualizing vector fields; however, the perceptual requirements for visualizing breathing motion are different and require alternative visualization approaches. Our preliminary user survey indicates the benefit of specific visualizations for specific perceptual problems and also showed the challenge in defining visualizations with a consistent perceptual experience for multiple users. In our work we also pointed out the lack of research in this specific medical visualization field. We believe that the future increase of publicly available 4D datasets will motivate further research in this field and introduce alternative approaches for handling these new and challenging visualization aspects.

References

1. Baker, S., Scharstein, D., Lewis, J.P., Roth, S., Black, M.J., Szeliski, R.: A database and evaluation methodology for optical flow. In: IEEE 11th International Conference on Computer Vision, ICCV, pp. 1–8. IEEE (2007)
2. Delhay, B., Clarysse, P., Pera, C., Magnin, I.E.: A spatio-temporal deformation model for dense motion estimation in periodic cardiac image sequences. In: Statistical Atlases to Personalized Models: Understanding Complex Diseases in Populations and Individuals, Satellite Workshop MICCAI 2006, pp. 87–90 (2006)
3. Engel, K., Hadwiger, M., Kniss, J., Rezk-Salama, C.: Real-time volume graphics. AK Peters Ltd. (2006)
4. Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense image registration through MRFs and efficient linear programming. Medical Image Analysis 12(6), 731–741 (2008)
5. Glocker, B., Sotiras, A., Komodakis, N., Paragios, N.: Deformable medical image registration: Setting the state of the art with discrete methods. Annual Review of Biomedical Engineering 13, 219–244 (2011)
6. Grady, L.: Random walks for image segmentation. IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI) 28(11), 1768–1783 (2006)
7. Handels, H., Werner, R., Schmidt, R., Frenzel, T., Lu, W., Low, D., Ehrhardt, J.: 4D medical image computing and visualization of lung tumor mobility in spatio-temporal CT image data. International Journal of Medical Informatics 76, S433–S439 (2007)
8. Honigmann, D., Ruisz, J., Haider, C.: Adaptive design of a global opacity transfer function for direct volume rendering of ultrasound data. In: Proceedings of the 14th IEEE Visualization 2003, VIS 2003, pp. 489–496 (2003)
9. Johnson, C.: Top scientific visualization research problems. IEEE Computer Graphics and Applications 24, 13–17 (2004)
10. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.: elastix: A toolbox for intensity-based medical image registration. IEEE Trans. Med. Imaging 29(1), 196–205 (2010)
11. Kosara, R., Healey, C.G., Interrante, V., Laidlaw, D.H., Ware, C.: Thoughts on user studies: Why, how, and when. IEEE Computer Graphics and Applications 23(4), 20–25 (2003)
12. Krüger, J., Westermann, R.: Acceleration techniques for GPU-based volume rendering. In: Proceedings of the 14th IEEE Visualization Conference, pp. 287–292 (2003)
13. Laramee, R.S., Hauser, H., Doleisch, H., Vrolijk, B., Post, F.H., Weiskopf, D.: The state of the art in flow visualization: Dense and texture-based techniques. Computer Graphics Forum 22(2), 203–221 (2004)
14. Post, F.H., Vrolijk, B., Hauser, H., Laramee, R.S., Doleisch, H.: Feature extraction and visualization of flow fields. Eurographics 2002 State-of-the-Art Reports 1, 69–100 (2002)
15. Prckovska, V., Peeters, T., van Almsick, M., Romeny, B., Vilanova i Bartroli, A.: Fused DTI/HARDI visualization. IEEE Transactions on Visualization and Computer Graphics 17, 1407–1419 (2011)
16. Sarrut, D., Boldea, V., Miguet, S., Ginestet, C.: Simulation of four-dimensional CT images from deformable registration between inhale and exhale breath-hold CT scans. Med. Phys. 33(3), 605–617 (2006)

17. Vandemeulebroucke, J., Sarrut, D., Clarysse, P.: The POPI-model, a point-validated pixel-based breathing thorax model. In: International Conference on the Use of Computers in Radiation Therapy (2007)
18. Yigitsoy, C., Rijkhorst, E., Navab, N., Wachinger, C.: Manifold learning for image-based breathing gating in ultrasound and MRI. Medical Image Analysis 16(4), 806–818 (2012)
19. Yigitsoy, M., Wachinger, C., Navab, N.: Temporal groupwise registration for motion modeling. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 648–659. Springer, Heidelberg (2011)

The Colored X-Rays

André Aichert*, Matthias Wieczorek, Jian Wang, Matthias Kreiser,
Lejing Wang, Pascal Fallavollita, and Nassir Navab

Chair for Computer Aided Medical Procedures (CAMP),
Technische Universität München, Munich, Germany
`andre.aichert@cs.tum.edu`

Abstract. Medical imaging has come a long way since the first X-Ray
pictures were taken in 1895 and X-Ray is still a primary image source for
diagnosis and intra-operative guidance in today's clinical setting. How-
ever, grayscale X-Ray images have some limitations, especially lacking
proper depth cues visible to the clinician. In the area of psychology, color
has been deemed as self-sufficient depth cues in the visual field. Thus,
the focus of this work is integrating color directly in X-Ray in order
to disambiguate the sequences of anatomy and hence provide intelligent
depth cues for quicker X-Ray interpretation. To achieve this we register
pre-operative CT data and X-Ray. A new transfer function is derived
using depth and intensity for color emission. Results from a question-
naire to surgeons and medical imaging experts and Likert scale show a
positive result of 4.4 average on a 5 scale. Furthermore, we assess the
impact of misalignment of the pre-operative CT data and show that the
color X-Ray image is very resilient to such errors.

1 Introduction

Little emphasis has been placed on the visualization of X-Ray images not only
in common practice but also in the research community. Even the well-known
first X-Ray image of a hand taken by Wilhelm Conrad Röntgen was presented
in exactly the same fashion as X-Rays in medical use today. In contrast, state-
of-the-art in visualization of CT and MR data has seen dramatic advancements.
Clinicians rely on X-Ray in daily diagnosis and intervention where it is very
common to take several X-Ray shots from perpendicular views, to better under-
stand anatomy. However, depth is difficult to perceive in single X-Ray images,
only the complicated mental fusion of two or more 2D X-Ray images enables a
recovery of 3D information. There is a lot of research in the computer graphics
community on occlusions, depth perception or focus and context visualizations
for volumetric data. For instance, in [2] volume rendering of CT is used to cut
part of the volume while maintaining contextual information, which leads to a
refined understanding of the 3D data and allows incorporation of ultrasound.
In [6], depth cues are added to volume renderings of 3D angiographic images to
support spatial comprehension and both depth perception of coloring, to achieve
a modified depth of field effect, and edge enhancement are evaluated in a small

---

* Corresponding author.

user study with promising results. The authors of [3] suggest color coding maximum intensity projection renderings of volumetric data. Their method translates depth directly into a specific color, based on where the maximum occurred. However, all of these works focus on 3D imaging. For 2D X-Ray, depth perception for purely absorptive lighting models using digitally reconstructed radiographs (DRR) has been studied in [5]. The authors show that humans are able to perceive depth in two X-Ray images through stereopsis. They also found that aerial perspective (i.e depth of field effect) can give strong depth cues. In the medical field this has been applied to 2D digital subtraction angiography (DSA) [10] but its use has not caught on in clinical practice - perhaps in part because it requires a change in hardware. We also observe that grayscale images are not an optimal form of presentation of medical data in terms of the human abilities to take up information.

Color-coding of funcional information in DSA sequences as presented in [7] is a promising effort to enrich DSA with clinically relevant and otherwise unused information. According to an interesting commentary [1] it is mostly up to the industry to pick up these concepts. In addition, we believe that clinical acceptance of other methods of post-processing and/or coloring for traditional modalities will also gain momentum if these efforts succeed in impacting clinical practice. In [9], a method is presented that relies on a two-pass digitally reconstructed radiograph (DRR) scheme to encode depth information from a CT and its registration to X-Ray. Their method fuses opaque geometry, such as surgical tools or a 2D ultrasound plane correctly with X-Ray and final results from a survey demonstrate promising results for depth perception.
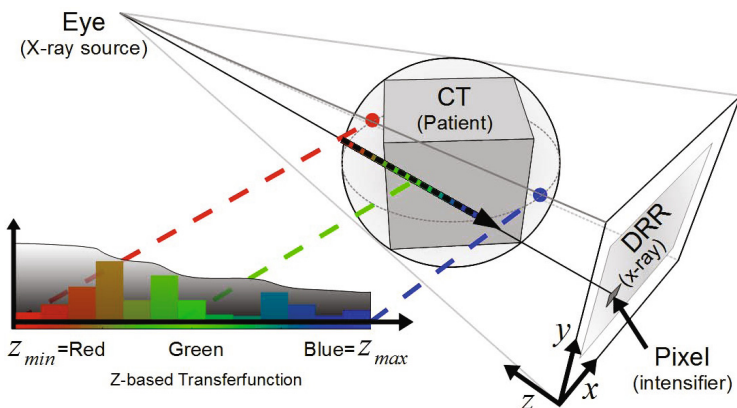

Fig. 1. Coloring based on depth in a typical DVR set-up in case of a registered CT and X-Ray

In this work we propose that depth perception of single X-Ray images can be significantly improved if additional depth information is encoded in color. In this way, the human sense of vision can be used to its full potential. Our method

aims at supplementing information through color without a negative impact on diagnostic value of the original X-Ray.

We begin from the basis that X-Ray is a purely absorptive modality. That means the intensity of a pixel goes back to absorption of a ray all the way through the patient, or CT volume respectively, indiscriminative of position. Depth can be seen as a distance in Z-direction to the X-Ray source. Similar to [3] we seek to encode depth in color. However, unlike in [3], there is no single depth value associated with a pixel in an X-Ray image. So the actual color of a pixel must be a continuous mix of colors, depending on how much absorption occurs in a particular region along the ray. Let us suppose that the depth information is provided by a 3D pre-operative or atlas data co-registered with the image.

2 Methodology

2.1 Colored X-Rays

We propose a volume rendering (VR) with a 2D depth-intensity-based transfer function (TF) that encodes depth in color during rendering of an emission-only model. To illustrate this intuition, imagine several layers parallel to the image plane of a luminescent gas with depth specific coloring. Although X-Ray is an absorption-only modality, images of an emission-only model are similar in the way intensity is accumulated indiscriminative of depth. To model this functionality we use a rainbow-like TF which maps depth to color, as illustrated in several figures. By convention, we choose the following color-depth encoding; red is used for absorption close to the observer, green for absorption around the center and blue for farther away absorption. Both historical research [8] and recent ones [4] agree with this choice, since red and green objects are perceived as closer than blue. However, this color scheme can easily be changed on the fly. For example, using just two colors for near and far may be easier to interpret in some cases. The final pixel color is accumulated along the ray (similar to the creation of a DRR). To illustrate the perceptive value of coloring, Fig. 2 shows a DRR of a torso compared to the same image colored according to depth. The figure also gives four qualitative examples of how emission color is directly related to the absorption as it occurs along the ray.

The color of a single voxel depends on depth and intensity and is therefore modeled as a mapping over a 2D domain. Since we aim at working with given X-Ray images, we decouple the DRR absorption model and the color emission model while maintaining a linear relation between absorption and emission. Thus we are not bound to a full 2D transfer function. Instead, we can model two 1D transfer functions - one for the mapping from depth to color and one for the mapping of the voxel values to the corresponding intensity, where the later is the same as in standard DRR generation.

$$tf : [0;1] \times [z_{min}, z_{max}] \mapsto [0;1]^3; (i, z) \mapsto \lambda(z) * \mu(i) \tag{1}$$

In Eq. (1) i is CT intensity and z is depth. The function $\lambda(z)$ maps depth in the viewing direction to an RGB color. $\mu(i)$ is the standard TF as in any DRR
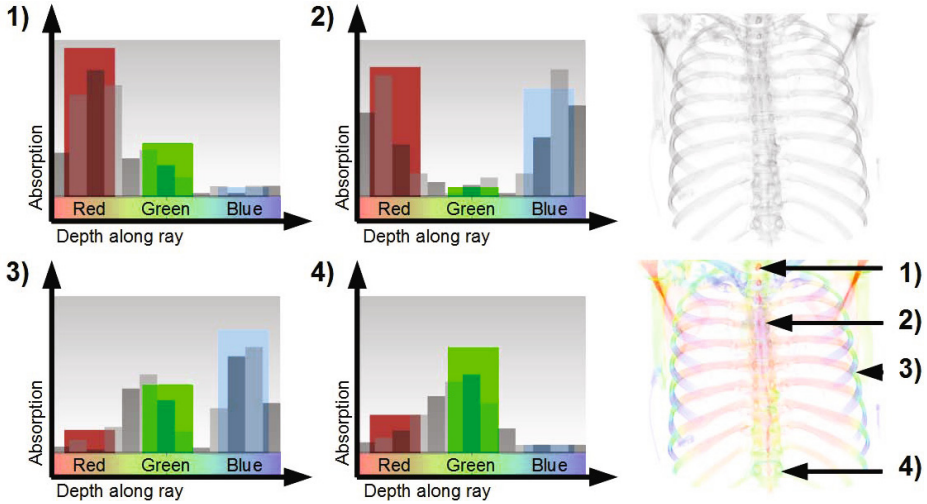
Fig. 2. 1) - 4) show the relation of pixel color to the distribution of absorption along the respective rays from the colored image. Note the color was added after the DRR above was rendered, not in the same process.

generation. The multiplication causes a weighted accumulation of the different colors along the ray. Such that high intensity values cause a higher contribution to the final color. For an optimal color contrast, we choose z_{min} and z_{max} tightly around the bounding box of CT (see Fig. 1). Thus only minimal portion of the color spectrum is wasted.

In many clinical cases an inverse camera model is used in combination with X-Ray based visualizations. This would simply flip the color ordering and would not affect the overall method. Actually, Fig. 3 shows a colored angiographic X-Ray using an inverse camera model. Just as in [6], the color makes it easy to distinguish between near and far structures. However, the coloring takes place directly in the X-Ray image. The catheter can be seen just as clearly as in the original X-Ray, wheras blending a volume rendered image on top of the X-Ray always loses some of the intensity information. Note that even in regions, where registration is off in the x-y plane, the assigned depth becomes clear as long as some of the vessel is colored, since the color also follows the shape of the structures it represents.
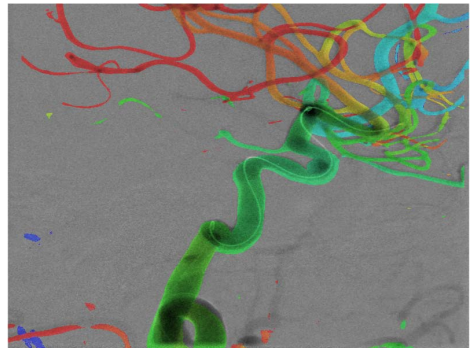

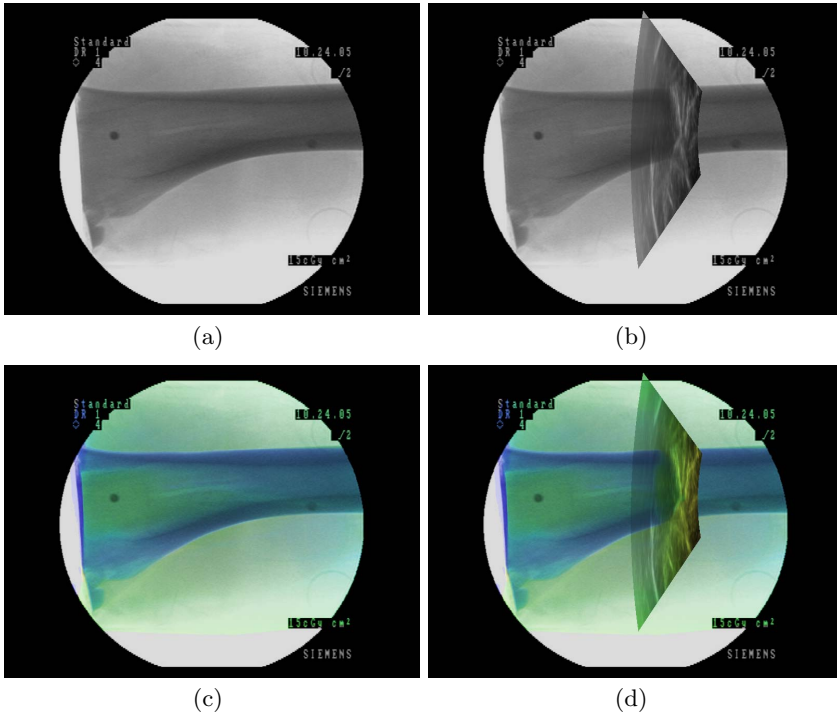**Fig. 3.** Colored angiographic X-Ray using inverse camera model

Fig. 4. (a) original X-Ray image, (b) Result of the X-Ray fused with an US using IXPV [9], (c) colored X-Ray, (d) our method combined with IXPV

We intend to manipulate existing X-Ray images and are therefore obligated to preserve or improve the original clinical value. The contribution of this paper is that we preserve the original image intensities, while augmenting the image with color information holding additional depth cues. RGB color space is not well suited for our problem since a cartesian representation of RGB color does not model human perception well. For example scaling a 3-vector of RGB not only changes brightness but also color tone and increasing the red channel makes the color look more red, but also brighter. There is a number of increasingly complex color models which are better suited. For our purposes, the HSL color space seems adequate, since it is a simple representation that holds color tone (hue), saturation and luminance (lightness) in separate channels. We use HSL space to transfer color by replacing the lightness component with luminance from X-Ray. We transfer color by taking the H and S components of the depth-colored image and merge it with the luminance of the X-Ray. As mentioned earlier, in [9] an *Interactive X-Ray Perceptual Visualization technique* (IXPV) is presented that aims at fusing instruments and ultrasound with X-Ray. This method relates to ours since both use a co-registered CT to manipulate X-Ray images. In fact, Fig. 4 and Fig. 5 show how our method can be integrated with IXPV in a straight-forward manner by adaptive depth color encoding of fused
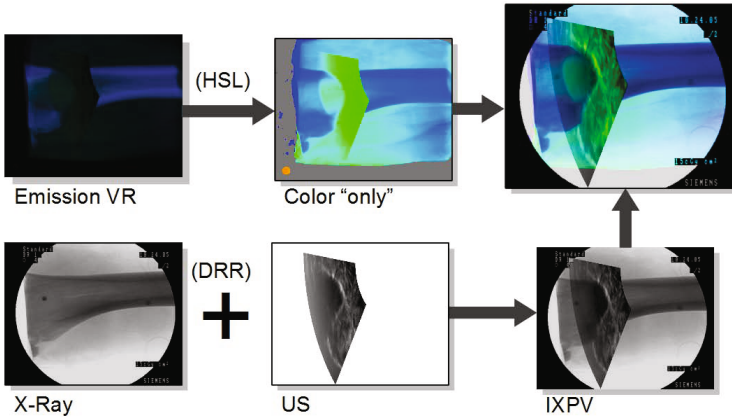
Fig. 5. Enhancing an X-Ray image based on information of CT data was also proposed in [9] and can be integrated with our method to color fused X-Ray and ultrasound images

X-Ray and ultrasound images. Additionally the comparison of IXPV, an X-Ray colored by our method, as well as a combination of both for a tracked US, fused with a colored X-Ray is shown. In case of the combination of both techniques the clipping of the DRR at the surface of the tracked instrument causes a shift in the coloring, such that only voxels in front of the instrument are used for the accumulation. This way, a movement of the instrument is also reflected intuitively in a color change, which provides a better understanding of both the position of the instrument and the 3D structure of the anatomy it intersects.

2.2   Effect of Registration Error

Since co-registered CT data of the same patient is not always available, our ultimate goal is to use non-patient specific CT data, optimally an atlas. Even when state-of-the art deformable registration were applied, this demands resiliency to a high degree of error in registration, due to potential differences between individuals and between atlas CT and patient's X-Ray image.

In Fig. 6 our method is applied to a high resolution X-Ray of a head. Notice that the front of the face and nose, as well as the spine are mostly green, wheres the skull, except at its base is violet. This relates to the fact that the brain tissue has a low attenuation coefficient as compared to the proximal and distal parts of the skull along the viewing direction. Consequentially the colors combined along the ray are primarily red (proximal skull region) and blue (distal skull region) resulting in purple. The nose and spine are on the central part of the ray and hence are colored in green. Notice the two white dots in the image. One is located on a red structure the other on a blue structure running in parallel. Due to coloring, they can be clearly assigned to front or back. Since the red structure is on the left, the patient must be looking slightly to the right. In this case the registration is off by as much as 2 mm in the x-y plane and the data

Fig. 6. Coloring applied to a high resolution X-Ray of a head. Note the improvement of perception in an area of fine structures despite large registration errors.

set depicts a region where depth perception is especially hard, due to identical fine structures from the right and left half of the base of the skull. Note that the perception is well present in spite of the existing errors in the registration. This is an important observation, because it demonstrates how even bad registration provides easily enough information to distinguish between right and left side of an object, for instance, or the relative order of structures in the field of view. Clinicians have to deduce depth of structures in traditional grayscale X-Ray image from their relation to macroscopic structures.

In Fig. 7, we validated the effect of registration error when blending an ultrasound plane directly in an X-Ray image. Depth errors were in the order of ± 5mm. Again, the ultrasound plane is correctly depicted in a green-tone color (i.e. closer to observer) whereas the cow tibia is shaded in blue color (i.e. farther away). Note the features from X-Ray such as the dark dot near the left-most red line are not affected at all by these errors. In areas where the ultrasound plane is present, error is expressed in a hardly noticeable shift in color (note the drawn lines as references). Finally, where the bone intersects the ultrasound slice the error in intensity from the IXPV method becomes visible. However, despite the error the general perception remains nearly the same for both images. Transferring only color information to X-Ray preserves intensity and therefore edges, which carry most information. Since no intensity information is taken from the CT, but only the color information, even completely wrong anatomy in the CT will have no effect on the image intensity. It is an advantage that even complete misalignment can never have negative effects on the intensity information a clinician can draw from the X-Ray. In [9] authors show that the absorption function approximation used to blend other objects with X-Ray becomes $\mu(z+\delta z) \approx \mu(z)$ when a translational error in z-component is introduced. Thus for their IXPV technique, such an error results in an offset within the grayscale intensity. For coloring X-Ray with our method, colors only provide relative visual information and the depth error will not affect the result. In essence, relative depth perception stays the same, since the ray goes all the way through the volume.
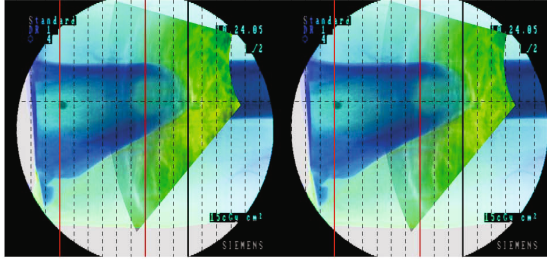
Fig. 7. Effect of registration error Tz in z-direction on color

3   Evaluation

Since we are addressing a perception issue, it is difficult to put hard numbers on the impact. We recall that the novelty of our method lies in the reintroduction of depth information from VR to an existing image, not image generation. We had hypothesized earlier that this approach makes a significant difference by improving the perception of classical x-ray imaging.

In order to evaluate the results of our color-based algorithm, we conducted an informal user study, based on a questionnaire, among a set of 11 people (i.e. five clinicians and six medical imaging experts). We accumulated statistical results based on the total number of correct answers from all participants. We asked only the surgeons to evaluate in a five-point Likert scale (i.e. 1 strongly disagree, 5 strongly agree) and asked whether they prefer the colored X-Ray images to the standard grayscale presentation. Participants were instructed to assess the influence of standard grayscale presentation and color-based depth en-

Fig. 8. The grayscale images are side-views that correspond to the colored top-view to their left

coding on two different synthetic datasets. Each consisted of a set of four helix and ring-like structures that was engineered to look similar seen from the top (Fig. 8 and Fig. 9). We used 2D static images of these volumes, showing lateral view, and the participants were asked to choose the correct match. To prevent learning (i.e. bias) we randomized sequence of images. Final results when considering all of the 11 participants average correct answers was 34 ± 21 percent for grayscale depth perception compared to 81 ± 19 percent for our color-based technique. Likert scale: the five surgeons responded positively to the color-based depth enhancement of X-Ray technique as suggested by a final average score of 4.4 ± 0.5 standard deviations.
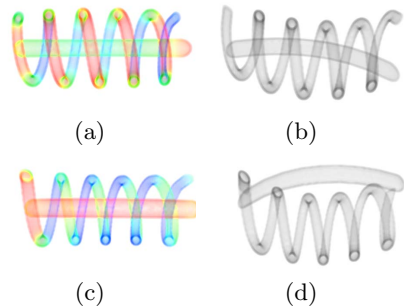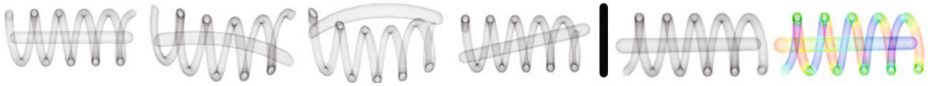
Fig. 9. Evaluation of depth perception of color in synthetic data. In this example, the four side views left of the bar would be presented along with only one of the views on the right (either grayscale or colored). The images on the right correspond to the structure just left of the bar (observed from the bottom).

4   Discussion and Conclusion

We propose a color-based visualization technique in order to improve depth perception in X-Ray. The use of depth color cues could provide an alternative to current medical practice when observing X-ray images. We reintroduce depth cues into X-Ray by transferring color hue and saturation of an emission based volume rendering of a CT where we encode the distribution of the absorption values in color. In spite of registration errors, the perception of the consecutive layers of anatomy within an X-Ray image is well preserved, which we hope will allow us to use a general anatomical atlas where patient-specific data is unavailable. We presented several example situations for different anatomy, including the head, the torso and an agiography of a cororary artery, and explained the advantages and gained information in each case. In addition, we showed that the generality and simplicity of the approach allows it to be combined into other systems, for example during fusion of ultrasound with single X-Ray images. From the questionnaire, clinicians received the colored X-Rays positively. More than 80 percent correct matches in the questionnaire were obtained. The introduction of color for encoding depth in X-Ray as well as for fusion of X-Ray and other modalities could change the way physicians use X-Ray imaging on a daily basis if this idea is further developed, accepted and validated.

References

1. Benndorf, G.: Color-coded digital subtraction angiography: the end of a monochromatic era? AJNR American Journal of Neuroradiology 31(5), 925–927 (2010)
2. Burns, M., Haidacher, M., Wein, W., Viola, I., Gröller, E.: Feature emphasis and contextual cutaways for multimodal medical visualization. In: Proceedings of Eurographics/IEEE VGTC Symposium on Visualization, EuroVis 2007, pp. 275–282 (May 2007)
3. Díaz, J., Vazquez, P.: Depth-enhanced maximum intensity projection. In: IEEE/EG International Symposium on Volume Graphics 8, pp. 1–8 (2010)
4. Guibal, C.R.C., Dresp, B.: Interaction of color and geometric cues in depth perception: When does ̈redmean ̈near: Psychological Research 69, 30–40 (2004)
5. Kersten, M., Stewart, J., Troje, N., Ellis, R.: Enhancing depth perception in translucent volumes. IEEE Transactions on Visualization and Computer Graphics 12, 1117–1124 (2006)

6. Ropinski, T., Steinicke, F., Hinrichs, K.H.: Visually supporting depth perception in angiography imaging. In: Butz, A., Fisher, B., Krüger, A., Olivier, P. (eds.) SG 2006. LNCS, vol. 4073, pp. 93–104. Springer, Heidelberg (2006)
7. Strother, C.M., Bender, F., Deuerling-Zheng, Y., Royalty, K., Pulfer, K.A., Baumgart, J., Zellerhoff, M., Aagaard-Kienitz, B., Niemann, D.B., Lindstrom, M.L.: Parametric color coding of digital subtraction angiography. AJNR American Journal of Neuroradiology 31(5), 919–924 (2010)
8. Verhoeff, F.H.: An optical illusion due to chromatic aberration. American Journal of Ophthalmology 11, 898–900 (1928)
9. Wieczorek, M., Aichert, A., Fallavollita, P., Kutter, O., Ahmadi, A., Wang, L., Navab, N.: Interactive 3D visualization of a single-view X-Ray image. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part I. LNCS, vol. 6891, pp. 73–80. Springer, Heidelberg (2011)
10. Worthington, C., Peters, T.M., Ethier, R., Melanson, D., Theron, J., Villemure, J.G., Olivier, A., Clark, J., Mawko, G.: Stereoscopic digital subtraction angiography in neuroradiologic assessment. AJNR Am. J. Neuroradiol. 6(5), 802–808 (1985)

Use of a Mixed-Reality System to Improve the Planning of Brain Tumour Resections: Preliminary Results

Kamyar Abhari, John S.H. Baxter, Elvis S. Chen, Ali R. Khan, Chris Wedlake, Terry M. Peters, Sandrine de Ribaupierre, and Roy Eagleson

Robarts Research Institute
Western University, London, ON, Canada

Abstract. The lack of intuitive visualization techniques for neurosurgical planning is a challenging hurdle faced by neurosurgeons and neurosurgery residents. Within this context, this paper describes the development and evaluation of an Augmented Reality (AR) system geared towards planning brain tumour resection interventions. Successful resection of a tumour or hematoma requires careful pre-operative planning to avoid damaging the brain. We hypothesize that our proposed AR system facilitates the planning of tumour resection operations by making more effective use of the visuospatial abilities of individuals to assess patient-specific data. To test our hypothesis, a number of experiments were conducted where subjects were asked to perform relevant spatial judgment tasks using three different conventional visualization approaches as well as the proposed AR system. Our preliminary results indicate that, compared to traditional methods, the proposed AR system a) greatly improves the user performance in tasks involving 3D spatial reasoning about the tumour relative to the anatomical context, b) reduces error associated with mental transformation, and c) supports generic spatial reasoning skills, independent of the sensory-motor tasks performed.

Keywords: Augmented Reality, Neurosurgical Planning, Tumour Resection.

1 Introduction and Clinical Motivation

Approximately 23,000 Americans and 2,800 Canadians were newly diagnosed with primary brain tumours in 2012, resulting in 13,700 and 1,800 deaths respectively [1][2]. Brain cancer with a five-year relative survival rate of 35% [1] is among the least survivable types of cancers. Thus, effective treatment is necessary. Compared to alternative courses of treatments, surgical resection of a tumour is the most recommended option [3]. Pre-operative planning of tumour resection interventions involves identifying optimal surgical pathways and specifying surgical entry points based on a number of criteria, geared towards minimizing post-surgery complications. Among the most important of which is the

avoidance of critical structures, such as functional areas of the brain, white matter tracts, and major arteries and veins. Additionally, determining the closest point on the skull to the target and aligning the surgical path with the tumour's longest axis, further reduces the damage to the brain. Formulating the optimal path and entry point, demands great spatial reasoning to cognitively process the inter-spatial relationships between the tumour and other structures and landmarks. Conventional visualization approaches to planning tumour resection interventions involve examining 2D orthogonal slices of pre-operative MR images, in order to estimate the target's shape, location, and size in the context of surrounding structures. The inherent limitations of standard 2D views and the reliance on spatial reasoning can be prone to error due to high cognitive load, and slow the process of planning [4]. On the other hand, orthographic and multiplanar reconstruction, segmented surface-based, and 3D volume rendering methods represent 3D data in a more usable fashion. However, they are limited by visualization and prohibitive interaction constraints that are associated with WIMP-based[1] applications [5]. While it is clear that users can be trained to perform well on non-intuitive tasks, they are more prone to errors when interacting with non-intuitive interfaces under additional cognitive demands. Although this is not controversial, the need for robust, intuitive visualization environments has remained largely unfulfilled for many neurosurgical planning applications.

When compared with classical interfaces, Augmented Reality (AR) environments provide intuitive modes of visualization and interaction [6]. AR environments make use of computer vision and computer-graphics techniques to visualize specific anatomical structures in the context of the scene captured from the observer's point-of-view using digital cameras. They also enable intuitive exploration and examination of 3D scenes through direct manipulation [7], making use of a manipulandum that mimics standard tools used in the OR. Natural gesture-based interaction with patient data reduces the cognitive load for navigation tasks, and yields rapid comprehension of complex spatial relations between anatomical structures. We believe that AR systems can facilitate the planning processes through specialized software and hardware that results in rapid, effortless, and accurate visualization and interaction. However, a number of perceptual ambiguities exist within AR visualization systems that lead to perceptual contention or ambiguities for the observer [8]. Our approach involves careful consideration of these constraints and the capacities of the human perceptual system during the design cycle. Using the proposed AR system, we hypothesize that planning of brain tumour resection intervention can be accomplished more accurately compared to traditional approaches. We note that the term 'accuracy' as used in this study refers to user accuracy independent of system accuracy, although both contribute to error. A number of experiments were conducted to test our hypothesis, and the usability of our AR system was evaluated in terms of reducing error associated with the user performance.

---

[1] WIMP stands for 'Windows-Icons-Menus-Pointer' and refers to a style of interaction using these elements of the user interface.

2   Materials and Methods

2.1   AR Environment

The AR system is comprised of a head-phantom and commercially available, off-the-shelf AR eyewear (Vuzix 920AR, Vuzix corporation, Rochester, NY), both of which are tracked using an optical tracking system (Polaris, NDI, Canada), (Figure 1). The Vuzix eyewear is equipped with twin cameras and displays to record and display stereo images. Tracking the head phantom and the Vuzix goggles allows for the correct fusion of virtual and physical spaces. Our visualization approach involves making use of a tracked stylus to provide a better visualization experience.
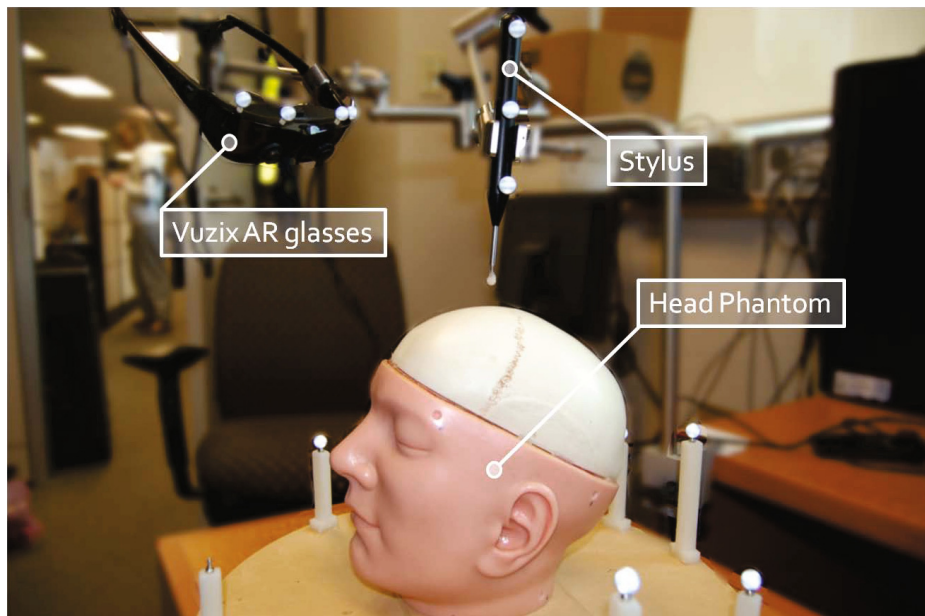

Fig. 1. Our AR platform includes Vuzix eyewear, a head phantom, and a tracked stylus

In a window-and-context paradigm [10], the tracked stylus was used to dynamically displace a focus window in which the volume of interest is visualized (Figure 2). Voxels that lie within this window are rendered using direct volume rendering (DVR) and two-dimensional transfer functions (2D TF), a standard approach for reducing the ambiguity and conveying subtle surface properties to the observer [12]. Furthermore, early ray termination is explicitly triggered through the use of an additional transfer function, which prevents rendering the volumetric data outside the focus window, known as context. This results in a keyhole-like aperture generated into the volume. Similar to virtual windows [18], the keyhole aperture would reduce depth misperception to some extent.

Such a technique also draws the attention to the region-of-interest while preventing cognitive overload (as information outside the region-of-interest is more likely to be ignored by observers[14]). Moreover, target structures can be delineated using volumetric Cel-shading to improve their visual perception. Cel-shading is a non-photorealistic technique to highlight the contour lines with thick black ink segments to emphasize edges. It is shown previously that enhancing the contours would improve the perception of continuity and depth [17].
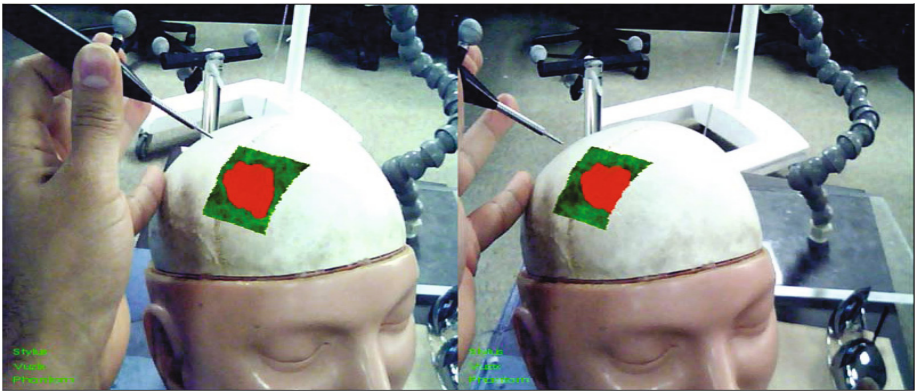

Fig. 2. To avoid misperception, yet visualizing patient data, a keyhole-like aperture is added to the stereo images of the AR system

The use of window-and-context visualization in AR environments is not a novel concept, as demonstrated by Bichlmeier et. al, who employed such a technique to improve depth perception in endoscopic procedures [18]. Their visualization approach, however, involved the use of segmented surface rendering (as opposed to direct volume rendering) to reach real-time rendering of the anatomical structures [18]. Compared to surface rendering, no explicit segmentation procedure is required to visualize different anatomical structures using DVR with two-dimensional TF. Moreover, volume rendering also provides a flexible approach to process translucent objects [19]. Despite its advantages, DVR has not received much attention as a visualization tool in AR environments because of its high computational cost and therefore low processing speed. In this study, however, we make use of a real-time (up to 40 fps for 512^3 dataset) volume rendering technique by taking advantage of specialized hardware technology and GPU programming [13]. The visualization consisted of a CUDA-accerelated (NVIDIA) Visualization Toolkit (VTK) pipeline.

Although visualization within our AR environment can provide a global appreciation of the target location and orientation with respect to the context, it is limited by the window size and therefore the amount of visualized information. Hence, we extended our interactive approach by incorporating a Virtual Reality (VR) mode into our visualization scheme. In this mode, virtual data can be visualized in its complete form by turning off the video feed and halting the early
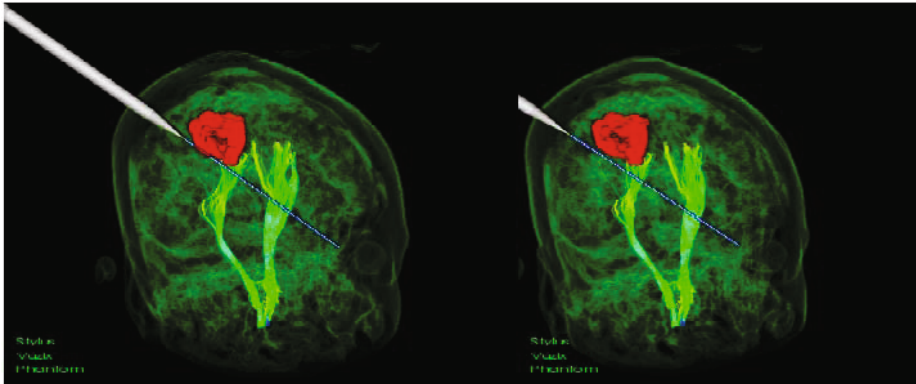
Fig. 3. Within virtual reality environment, additional information, such as diffusion tensor imaging (DTI) tracts or virtual representation of the stylus and its trajectory, can be visualized with minimum visual clutter

ray termination process (Figure 3). Virtual reality can be used to effectively visualize additional information, such as DTI tracts or a virtual representation of the stylus and its trajectory, when optimized to reduce visual clutter. While the AR mode provides a better understanding of the spatial relationship between the virtual data and the physical context, visualization of too much information through the key-hole aperture would lead to visual clutter and thus, increase the cognitive load. On the other hand, the VR mode is suitable for visualizing more information at the cost of reducing contexual appreciation. We may benefit from both the AR and VR by toggling between these two modes of visualization. A foot pedal was provided to participants allowing them to switch back and forth between these two modes.

2.2   Experiments

The objective of our evaluation process was to compare the user performance associated with conventional approaches against our AR system. Conventional techniques consist of two-dimensional slice-by-slice (2D), crossed-plane (XP), and three-dimensional volume rendering (3D) 4. Ten novice subjects (8 Male, 2 Female, all graduate students), participated in our validation study and were asked to perform a series of relevant spatial reasoning tasks while exploring data via available techniques. These tasks were defined based on the planning principles mentioned earlier in the introduction. To better understand the contributing factors, each task isolates one particular criterion as follows: 1) finding the maximal distance from the target to a nearby critical structure, and thereby avoiding it; 2) finding the shortest distance to the target from the surface of skull; and 3) determining the longest axis of the target. Based on these criteria, subjects were asked to make use of the head phantom and indicate the optimal entry point/surgical path by placing the tip of the stylus on their chosen entry point while aligning it with their chosen surgical path.
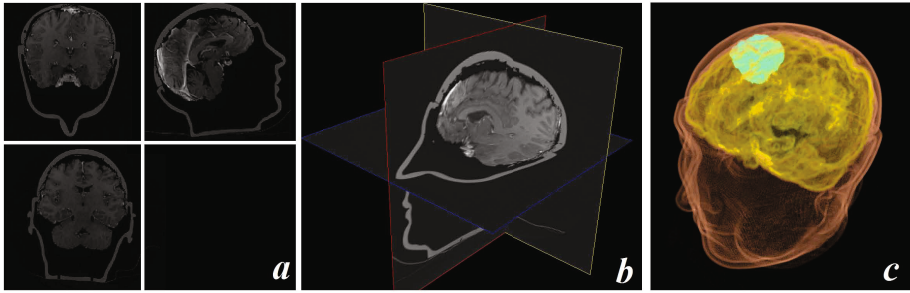
Fig. 4. Conventional visualization approaches involve two-dimensional a) slice-by-slice (2D), b) crossed-plane (XP), and c) three-dimensional direct volume rendering (3D)

Stimuli consisted of the head phantom's CT images which were modified to meet the requisites of each specific task (Table 1). Phantom data designed to mimic realistic scenarios was used to increase the ecological validity, while target structures were simulated to increase our control over the experimental design. Simulated structures were analogous to patient anatomical data. Experiments were conducted in two different phases. In the first phase:
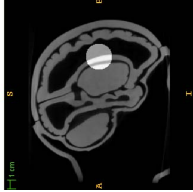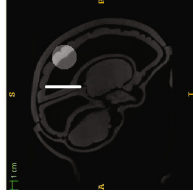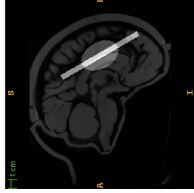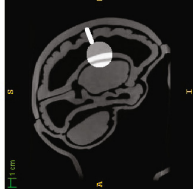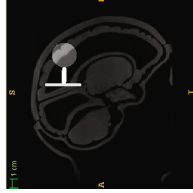
1. Synthetic ellipsoids were added to the phantom's CT image in a pseudo-random fashion to simulate tumours. The task was defined as finding the longest axis of the ellipsoid. The ground truth vector was computed using principle component analysis (PCA). This task was designed to isolate the longest axis criterion.
2. Synthetic spheres were added to the CT volume in a pseudo-random fashion to simulate tumours. The task was defined as finding the shortest distance from the surface of the tumour to the phantom's skull. The ground truth was determined by segmenting the skull[2] and finding the shortest possible distance. This task was designed to isolate the shortest distance criterion.
3. Synthetic spheres and tubes were added to the CT volume in a pseudo-random fashion to simulate tumours and critical pathways such as DTI tracts. The task was specified as finding the maximal distance between the surface of the sphere and the tube. The ground truth was determined by finding the Hausdorff distance between two surfaces. This task was designed to isolate the avoidance criterion.

In the second phase, subjects were asked to perform the same tasks as the first phase, but in this stage, the ground truth entry points and surgical paths were provided by adding visual assistance: the longest axes, the shortest paths, and the maximal distances were illustrated via synthetic lines (Table 1).

The rationale for this experimental design was, 1) to study the effect of different visualization approaches on user performance, and 2) to determine whether

---

[2] We used a segmentation interface described in [20].

Table 1. Stimuli and their corresponding tasks

Stimuli	Longest Axis	Shortest Distance	Maximal Distance
Task	Finding the longest axis of the target	Finding the shortest distance from target to skull	Finding the maximal Hausdorff distance between target and synthetic tube
Method of determining ground truth	Determined by primary component from PCA	Determined by segmenting skull and running kd-tree search	Determined by running kd-tree search
Sample stimulus in 2D (Phase 1)			
Sample stimulus in 2D (Phase 2)			
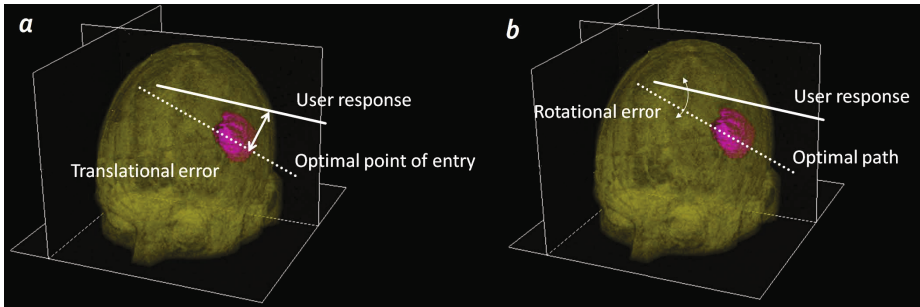
Fig. 5. Rotational and translational error were used as metrics to measure users' performance. Rotational error is measured as the deviation (in degrees) between the chosen and optimal paths. Translational error is measured as the Euclidean distance (in mm) between the optimal and chosen points of entry.

providing visual assistance can diminish the difference between available visualization techniques. User performance was measured based on the translational and rotational error values. Translational error is measured as the Euclidean distance (in mm) between the optimal points of entry and those selected by participants using the stylus (Figure 5-a). Similarly, rotational error is measured as the deviation (in degrees) between the chosen and optimal paths. (Figure 5-b). The optimal entry point and the optimal path were determined based on the particular criterion isolated for the particular task.

3   Results and Discussion

Each experiment involved 12 trials (3 tasks x 4 visualization methods) where participants ($n = 10$, no prior training) were presented with a randomized collection of synthetic phantom data. To minimize the effects of learning and fatigue, the stimuli (CT images), the visualization mode, and the task were all randomized and counterbalanced. The user overall performance, as shown in Figure 6, was calculated by averaging the rotational and translational error over each of the visualization techniques. The same process was repeated for the second phase where the ground truth was provided visually to the user as a synthetic tube (Table 1).
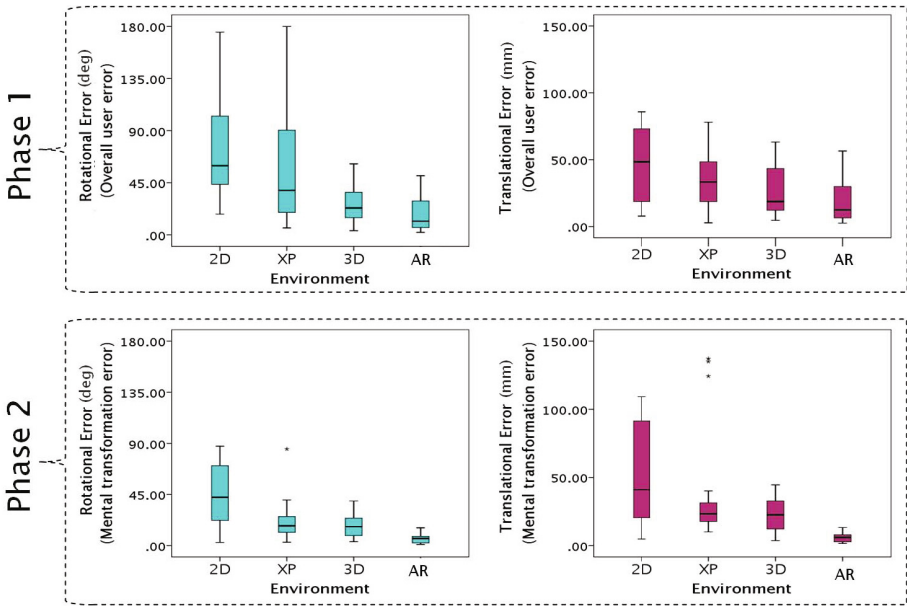
Fig. 6. Overall rotational and translational error values observed in phase 1 (top) and phase 2 (bottom)

Phase 1 (Without Visual Assistance). A multivariate ANOVA test indicated that the mode of visualization was indeed significant in the first phase (rotational error: $F=13.37$, $p<0.005$, translational error: $F=8.11$, $p<0.005$). This level of significance was derived from the Šidàk correction: $\beta = 1 - \sqrt{1 - 0.01} \approx 0.005$, which would lead to a combined level of signifance of 1%. No interaction effect between visualization method and task was observed. Post-hoc analysis using Tukey HSD test indicated that the difference between 2D/XP visualization environments and 3D/AR were statistically significant (Figure 7). However, the difference between the 3D and AR environments was not statistically significant.
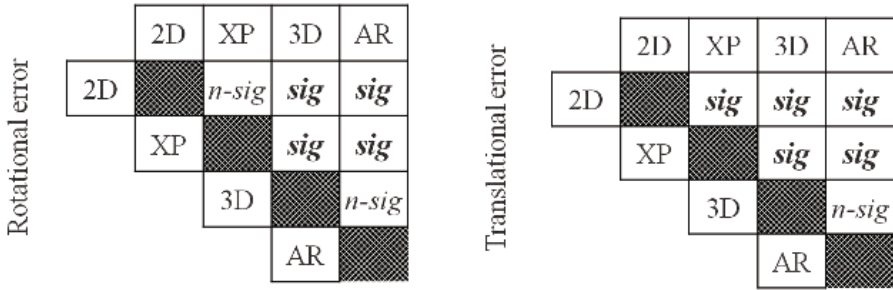
Fig. 7. Results of the Tukey HSD post-hoc analysis (sig: $p < 0.01$, n-sig: $p > 0.01$)

Therefore, it can be concluded that when there is no visual assistance, a) the method of visualization significantly affects the user performance error, and b) the impact of method of visualization is not affected by the task performed (and vice versa). This indicates the generic usability of the visualization regardless of the task performed. Therefore, the information is available for untrained and generic perceptual usage.

Phase 2 (With Visual Assistance). Unlike the previous stage, significant interaction among factors was observed (Figure 8), i.e. the magnitude of difference between the means in environments depends in part upon the task performed. Therefore, a test of significance must be conducted for each visualization method under each task performed. Performing such a test reveals that most interactions occur in 2D, XP, and 3D environments (as illustrated in figure 8) [3]. This indicates the lack of generic usability of such environments. Additionally, the large variation of the rotational and translational error under various tasks in 2D and XP illustrates the veridicality of 3D and particularly AR environments.

Discussion. The spatial reasoning tasks involved in our experiments can be decomposed into two main components: 1. Spatial perception to locate the targets and judge their orientation, and 2. Mental transformation from 2D displays into 3D context. In a clinical sense, mental transformation takes place when the information collected from 2D monitors during the planning process is translated into the patient and operation environment. Significant improvement of performance in AR/3D in the first phase demonstrates that 3D perception of the target location/orientation in 2D/XP, and reasoning about the spatial relationship with the anatomical context, can be facilitated with appropriate visualization methods. Significant improvement of performance in AR while providing visual assistance illustrates that the AR mode of presentation reduces the need for mental transformation needed in the 2D/XP/3D viewing modes, thus reducing

---

[3] The estimated marginal mean values and the mean values are equivalent due to equal sample sizes in each group.
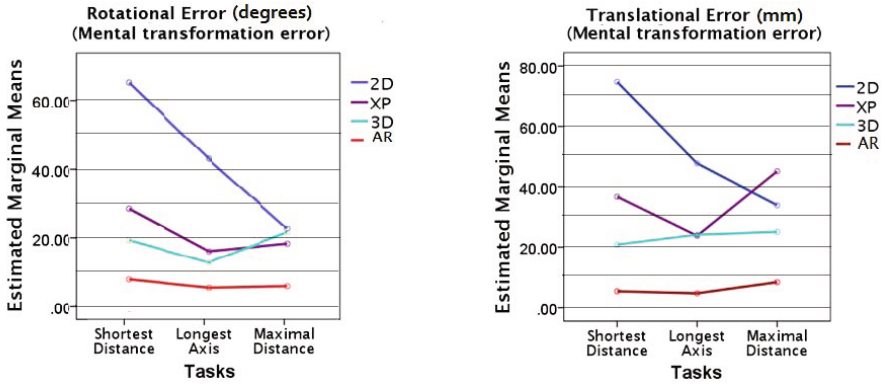
Fig. 8. Interaction in between spatial tasks and conventional visualization approaches (2D/XP/3D) is evident

user dependent error. By fusing the virtual space of the data and the physical spaces of the patient and the tool, we believe that AR contextualizes the tool within the display, making the tool affordances far more accessible and intuitive. Furthermore, insignificant changes of error in AR/3D under various tasks indicates the generic usability of the view provided by such environments, suggesting its use for tasks with multiple and possibly conflicting criteria.

4   Conclusion

Locating relevant structures such as tumours and tracts and perceiving the spatial relationships between them is necessary for successful neurosurgical preoperative planning. The usability of such percepts is heavily influenced by the mode of visualization and interaction within the neurosurgical planning environment. In this study, we investigated whether performing relevant spatial tasks could be facilitated by visualizing MR images using direct volume rendering within an AR environment. Accordingly, we conducted a number of experiments where subjects were asked to perform relevant spatial judgment tasks using four different visualization approaches. Our preliminary results indicate that, compared to traditional methods, AR environments could improve the spatial reasoning of subjects with respect to clinically-relevant criteria, leading to the conclusion that the planning of brain tumour resection interventions can be facilitated by the use of this visualization methodology. In particular, AR greatly improves the user performance in targeting shown by reduced errors associated with each task. Performance in the AR environment is more independent from the specific spatial tasks indicating that the visual information is more usable for generic tasks. As part of future work, we plan to extend our study by 1) recruiting clinicians (with different levels of expertise) as subjects; 2) using more clinically-relevant dataset; 3) improving the visualization approach; and 4) evaluating the efficiency of the AR environment. Nevertheless, the proposed work

has the potential to promote the quality of tumour resection planning and similar operations by considering the cognitive and perceptual limitations of human observers.

References

1. American Cancer Society: Cancer Facts and Figures 2012. American Cancer Society, Atlanta (2012) (last accessed December 12, 2012)
2. Canadian Cancer Statistics, produced by Canadian Cancer Society, Statistics Canada, Provincial/Territorial Cancer Registries, Public Health Agency of Canada (2012) (last accessed December 12, 2012)
3. National Cancer Institute (NCI) booklet (NIH Publication No. 09-1558), (Posted: April 29, 2009)
4. Hegarty, M., Keehner, M., Cohen, C., Montello, D.R., Lippa, Y., Allen, G.L.: The role of spatial cognition in medicine: Applications for selecting and training professionals. In: Applied Spatial Cognition: From Research to Cognitive Technology, pp. 285–315. Lawrence Erlbaum Associates Publishers (2007)
5. Biström, J., Cogliati, A., Rouhiainen, K.: Post-WIMP User Interface Model for 3D Web Applications, Helsinki University of Technology Telecommunications Software and Multimedia Laboratory (2005)
6. Traub, J., Stefan, P., Heining, S.M., Sielhorst, T., Riquarts, C., Euler, E., Navab, N.: Hybrid navigation interface for orthopedic and trauma surgery. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 373–380. Springer, Heidelberg (2006)
7. Hutchins, E.L., Hollan, J.D., Norman, D.A.: Direct Manipulation Interfaces. In: Norman, D.A., Draper, S.W. (eds.) User-Centered System Design, pp. 87–124. Lawrence Erlbaum, Hillsdale (1986)
8. Sielhorst, T., Feuerstein, Navab, N.: Advanced medical displays: A literature review of augmented reality. J. of Display Tech. 44, 451–467 (2008)
9. Prastawa, M., Bullitt, E., Gerig, G.: Simulation of Brain Tumors in MR Images for Evaluation of Segmentation Efficacy. Medical Image Analysis (MedIA) 13(2), 297–311 (2009)
10. Kruger, J., Schneider, J., Westermann, R.: ClearView: An Interactive Context Preserving Hotspot Visualization Technique. IEEE Transactions on Visualization and Computer Graphics 12(5), 941–948 (2006)
11. Drebin, R.A., Carpenter, L., Hanrahan, P.: Volume rendering. SIGGRAPH Comput. Graph. 22(4), 65–74 (1988)

12. Kniss, J., Kindlmann, G., Hansen, C.: Interactive volume rendering using multi-dimensional transfer functions and direct manipulation widgets. In: Proc. IEEE Visualization, pp. 255–262 (2001)
13. Baxter, J.S.H., Peters, T.M., Chen, E.C.S.: A unified framework for voxel classification and triangulation. In: Proc. SPIE, vol. 7964, p. 796436 (2011)
14. Williams, L.J.: Tunnel vision induced by a foveal load manipulation. The J. of the Human Factors 27(2), 221–227 (1985)
15. Lerotic, M., Chung, A.J., Mylonas, G., Yang, G.Z.: *pq*-space based non-photorealistic rendering for augmented reality. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 102–109. Springer, Heidelberg (2007)
16. Interrante, V., Fuchs, H., Pizer, S.M.: Conveying the 3D Shape of Smoothly Curving Transparent Surfaces via Texture. IEEE Trans. Visualization and Computer Graphics 3(2), 98–117 (1997)
17. Abhari, K., Baxter, J.S.H., de Ribaupierre, S., Peters, T., Eagleson, R.: Perceptual Improvement of Volume-Rendered MR Angiography Data using a Contour enhancement Technique. In: International Society for Optics and Photonics (SPIE), USA, vol. 8318, p. 831809 (2012)
18. Bichlmeier, C., Wimmer, F., Michael, H.S., Nassir, N.: Contextual anatomic mimesis: Hybrid in-situ visualization method for improving multi-sensory depth perception in medical augmented reality. In: Proc. IEEE and ACM Int. Symp. on Mixed and Augmented Reality, ISMAR, pp. 129–138 (2007)
19. Zhang, Q., Eagleson, R., Peters, T.M.: Volume visualization: A technical overview with a focus on medical applications. J. Digit. Imag. 24(4), 640–664 (2011)
20. Rajchl, M., Yuan, J., Ukwatta, E., Peters, T.M.: Fast Interactive Multi-region Cardiac Segmentation With Linearly Ordered Labels. In: 9th IEEE International Symposium on Biomedical Imaging, ISBI (2012)

Development of New Augmented Reality Function Using Intraperitoneal Multi-view Camera

Naoki Suzuki and Asaki Hattori

Institute for High Dimensional Medical Imaging,
The Jikei Univ. School of Medicine, Japan
{nsuzuki,hat}@jikei.ac.jp

Abstract. We developed a multi-view video camera system that works inside the abdominal cavity to obtain a wider range of information during laparoscopic and robotic surgery. This video camera system is able to enter the abdominal cavity and surround an organ in a fixed degree. We conducted in vitro and in vivo experiments to clarify the functions of the system. The advantage of the system is that it can alter the viewpoint without physically moving the camera and can obtain diversified intraperitoneal information. Moreover, by taking advantage of its video image array with geometrical regularity, we build a novel augmented reality function compared with the images for conventional navigation surgery.

Keywords: Multiple-viewpoint camera, Laparoscopy.

1 Introduction

Although more than 20 years have passed since laparoscopic surgery was first developed, intraperitoneal images have not undergone much change. Image quality has improved due to the spread of high-definition video images, but there are still various restrictions to images captured by the lens at the tip of a long tubular laparoscope. This is because the laparoscope itself has a limited field of vision and restriction in the direction in which it can see [1], [2]. As a result, the surgeon must operate in a restricted and narrow field of view. This may lead to the surgeon to not notice bleeding. Intraperitoneal accidents, in which the surgeon damages soft tissue with a surgical instrument because of a limited operative view, occur even to this date. For the same reason, second surgeries due to overlooking ruptured sutures and ligation failure have been reported [3-5]. This is one reason why we did not obtain intraperitoneal image information from a laparoscope. Instead, we positioned a group of cameras to obtain intraperitoneal image information. By taking this approach, we attempted to formulate a new augmented reality that allows for utilization of the attributions of image information. We verified of the function with extracted organ and animal experiments. We herein report on our development.

2   Method

2.1   Structure of Multi-faceted Camera and Prototype

A laparoscope's field of view is limited because it can only be set intraperitoneally in a conic shape with the trocar in the center [1]. The space inside the abdominal cavity can be widened by insufflation, but the organ and the tip of the laparoscope are sometimes too close, which results in a restricted view of certain parts of the organ. Our concept is to acquire views from multiple directions at the same time by positioning geometrically aligned small-size video cameras. Preceding studies and patents have captured operational field images from multiple viewpoints by attaching multiple cameras at the tip of the endoscope [6], [7], but these do not have structures that can film target organs from multiple directions.

   We aimed to deploy a new augmented reality function that obtains multiple types of information. This is in addition to geometrically aligned camera groups that obtain various viewpoints without physically moving during laparoscopic surgery. To realize this concept, we developed a group of small-size cameras that are positioned in a fixed degree surrounding the target organ in an arc. We previously developed a method that positions 60 video cameras in a circle to analyze human full body motion [6-8]. Although the size differs, we aimed to apply the method obtained from this past experiment to the present multi-view camera. We decided to position 5 to 8 cameras in the abdominal cavity considering the size of current video camera devices.

   It is impossible set up a group of multi-view cameras positioned on an arc with a radius of 50 mm or more through a trocar with a 10-mm diameter. Therefore, the camera assembly was divided into left and right parts and has a collapsible structure as shown in Figure 1. It is transformed into a bar-like shape to go through the trocar and reach the abdominal cavity. Figure 1 shows its basic structure. The camera group can be repositioned into an arc inside the abdominal cavity by pulling wires outside the body through the trocar. It has a geometrical structure so that each unit mounted with a small-size camera element links with a unit next to it by the pull of the wire to make an arc. After surgery, the surgeon loosens the wires for deployment and pulls them to fold the camera group and return it to its original bar-like shape so that it can exit the abdominal cavity through the trocar.
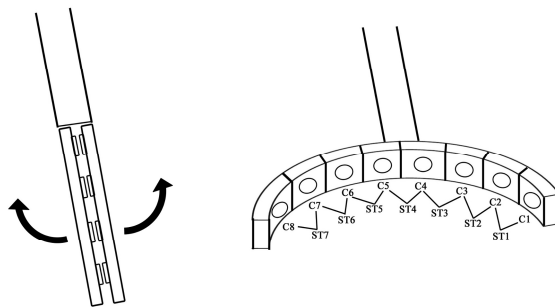

Fig. 1. Basic structure of multi-view camera

The prototype for the animal experiment comprised a multi-view camera that mounted video camera elements (maximum resolution, 1280×960) (Figure 2) on 50- to 70-mm-radius arcs. Electric circuits to drive each video camera element were collected onto a substrate. Light sources for each camera could be positioned on the substrate, providing even lighting for the target organ.

Wiring from the substrate in the back of each camera assembly by a flexible cable reduced the volume of cables for the electrical power source and signal lines to the substrate. This helped to miniaturize the camera assembly in the abdominal cavity. Accuracy of the device can be maintained through calibration inside the body after it is deployed even when geometric error occurs in camera array due to loosening of the wires that make up the arc or by soft tissue fragments being caught between movable parts.
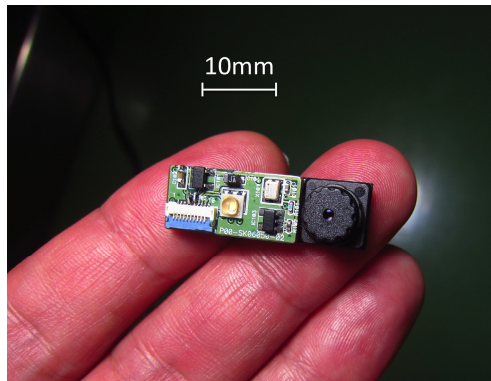
Fig. 2. Form and size of camera module used in prototype system

2.2   New Visual Information and Formulation of Less Restricted Augmented Reality

Obtaining Free Viewpoint in the Abdominal Cavity. By choosing viewpoints from 8 video cameras positioned in an arc when viewing a target object intraperitoneally, the surgeon is able to change the viewpoint without physically moving the camera inside the abdominal cavity. Because each camera drives independently, the surgeon can also monitor all of the viewpoints to understand the entire situation inside the abdominal cavity from various directions by multiple monitors. Moreover, by taking advantage of viewpoints of cameras situated next to each other, stereoscopic viewing is possible. The surgeon is able to choose stereoscopic images randomly from the n-1 direction versus a number of camera viewpoints (n). The surgeon can also obtain stereoscopic view without physically moving the cameras.

Superimposed Display of Organ Structure Using Augmented Reality. This augmented reality function enables the surgeon to recognize structures under the organ surface that cannot be seen grossly by superimposing structures of vessels and tumors in the target organ obtained by X-ray CT and MRI before surgery onto intraperitoneal images obtained from multi-view cameras. Thus, it is also equipped with a navigational

surgery function. Moreover, this function takes advantage of a free choice of viewpoint from multi-view cameras. It is able to superimpose images without physical movement of the cameras. During image processing, 2 cameras take in image signals continuously by a switcher to realize a stereo image function. Using these 2 time-sequential images from neighboring viewpoints, stereo-paired overlaid images of intra-organ structures are produced. That is, real-time stereo-paired overlaid images for the left and right eyes are produced by compositing surface-rendering models of vessels and tumors that are simultaneously reconstructed according to each viewpoint. Figure 3 shows the device composition of this function.
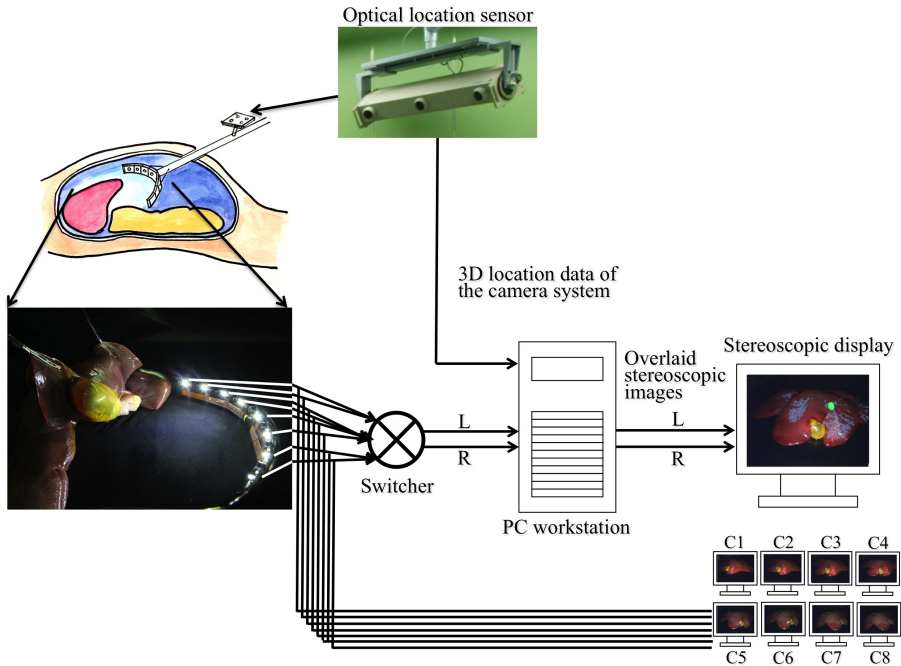
Fig. 3. Device composition of the system

3   Results

3.1   Obtaining Free Viewpoint and Stereoscopic Images

The figures show images taken from viewpoints of 8 directions using extracted liver and gall bladder. Figure 4 shows the setup of the experiment. Figure 5 shows images obtained from each viewpoint. The position of the image matches the number on the camera. As a result, we were able obtain 640×480 images in 30 frames/sec from 8 points in a 160-degree direction surrounding the target part. This confirmed that we were able to observe surgical organ conditions with a less restricted view without physically moving the camera inside. It also displayed viewpoints on the main monitor
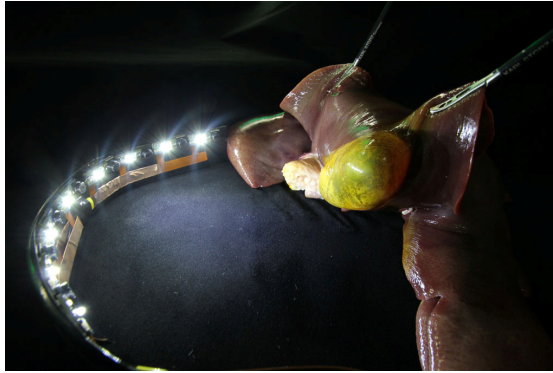
Fig. 4. In vitro experiment using extracted pig liver and gall bladder
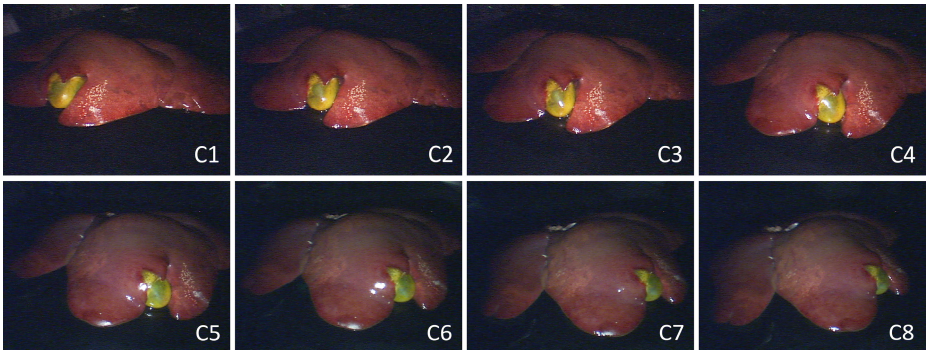
Fig. 5. Images obtained from each viewpoint when the camera system surrounded the extracted pig liver and gall bladder
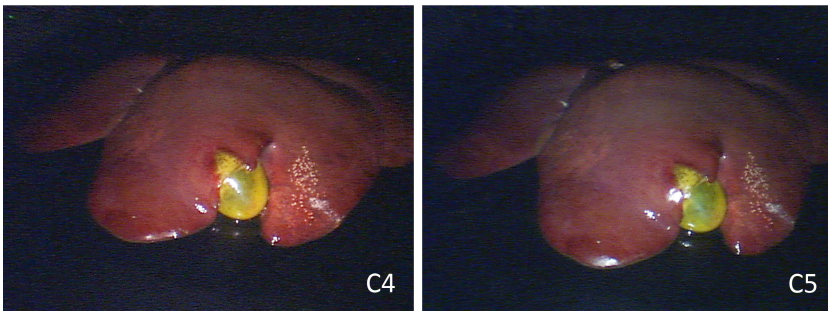
Fig. 6. Stereoscopic images from a viewpoint when the camera system surrounded the extracted pig liver and gall bladder

arbitrarily selected by the control unit. We enabled images from all 8 monitors according to the needs of the surgeon. These were able to be displayed on the submonitors. Figure 6 shows stereoscopic images chosen from 2 camera images next to each other; they were used as the right- and left-eye views, respectively.
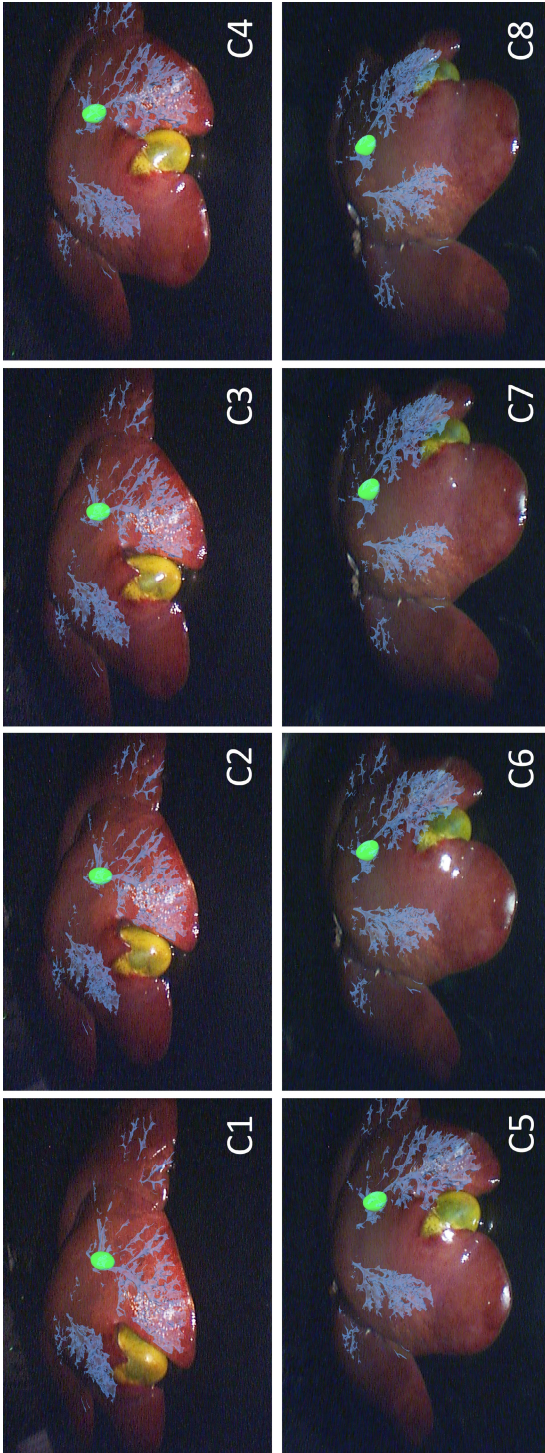
Fig. 7. Superimposed display of images obtained from each viewpoint of part of the vessels and artificial tumor inside extracted pig liver

3.2   Superimposed Display of Inner Organ Structure Using Augmented Reality

Figure 7 shows a superimposed display of intraperitoneal images using a surface model by carrying out segmentation of a part of the vessels and artificial tumor inside a liver from CT datasets obtained in advance with extracted organs of a pig. By utilizing the characteristics of the system, we were able obtain superimposed display images of the position and shape of the vessels in the organ and artificial tumor without physically moving the camera. We also displayed images from all 8 cameras according to the needs of the surgeon on submonitors, and we studied their applications. Figure 8 shows stereoscopic images obtained from cameras C4 and C5, situated near the center of the camera assembly arc. They show the surgeon bringing laparoscopic surgical forceps close to the artificial tumor depending on the superimposed display images. The stereoscopic images were taken at a frame rate of 26 to 29 frames/sec. They were displayed by delay of several frames. Although the frame rate slowed to 1/4 when superimposed images were displayed from 7 directions, we were able to display them simultaneously.
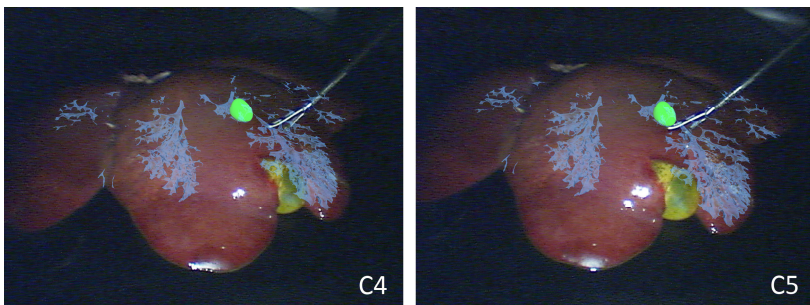
Fig. 8. Stereoscopic images from adjacent cameras (cameras 4 and 5) showing part of the vessels and artificial tumor inside extracted pig liver. Surgeons bringing laparoscopic surgical forceps close to the artificial tumor depend on these superimposed display images.

3.3   Obtaining Intraperitoneal Images from Animal Experiment

We conducted an in vivo verification experiment of the function of the system. Using 4 pigs weighing 35 to 40 kg, we inserted this system into the abdominal cavity under anesthetic conditions. Figure 9 shows image groups of the upper part of the abdomen (superior direction) from 5 viewpoints taken from the bottom part of the abdomen. Each image is provided at 30 frames/sec to the operator. Figure 10 shows the forceps being used from each viewpoint.

Fig. 9. Image groups taken from the bottom part to the upper part (head region) of the abdominal cavity from five viewpoints. Images are from pigs under anesthetic conditions in *vivo*.
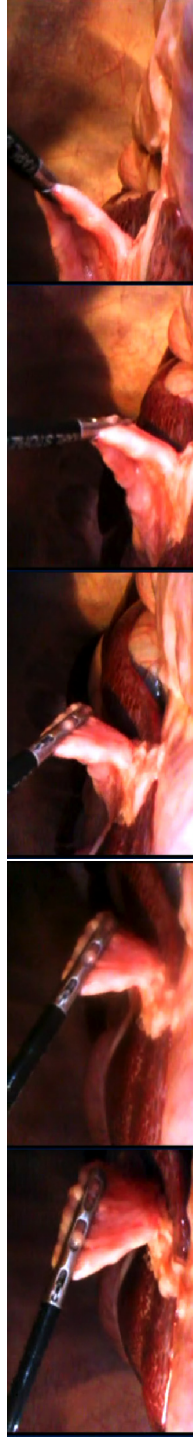
Fig. 10. Images of forceps being used in the abdominal cavity in a pig under anesthetic conditions from various viewpoints

4   Conclusion

We developed a prototype of a camera size that can be used in the abdominal cavity. We evaluated the system using isolated organs and animal experiments. In the in vitro experiment using extracted liver with gall bladder, we confirmed information obtained from the developed system. We were also able to evaluate the application of image information obtained from cameras surrounding the target parts.

The most important function of this system is that it can change viewpoints without physically moving the camera. This prevents clashes of laparoscopes and soft tissue in areas where conventional laparoscopes cannot see and achieves safe camera maneuvering. Conventional laparoscope images are limited to panning of the camera in a cone that centers on the trocar and limits sweet spots, significantly restricting optimum viewpoints. This system views the inner landscape from the target part, which enables views of various sweet spots. It is also able to view target parts from various viewpoints, which allows for different kinds of observations that are not possible using conventional methods. Thus, the surgeon and assistant can respectively choose the viewpoints they wish to see on the monitor. Moreover, they can see all viewpoints from all directions simultaneously using submonitors. The system also has a stereoscopic view function. They can also choose observation points from stereoscopic images from the n-1 direction. These functions, which enable operators to choose their viewpoints in the abdominal cavity, can provide vision information that may lead to new types of surgery. Functions of inner organ structure that are displayed by augmented reality and superimposed displays of planning data may also provide new visual information. We believe that these superimposed displays from various viewpoints without moving the cameras will increase the safety and speed of operations. By creating a prototype and evaluating its functions, we demonstrated that we are able to bring new viewpoints to laparoscopic surgery by obtaining unlimited viewpoints without physically moving the cameras. We also confirmed that adding depth to the field of view enables a safer operation.

We plan to develop a function that enables three-dimensional construction of the operational field by capturing continuous images of the field in a determined space of time. We are also considering mounting a feature inside the system that will update 4-dimensional information of the change in the operational field. This will be of use in evaluating accidents and evaluating the skill level of the operators. We plan to enhance the function of the control unit and create a system that can be used for clinical testing.

We think the system can be applied for clinical application in surgery in wide areas of the abdominal region including conventional laparoscopic surgery in the liver, pancreas, and colon. We also think we can provide various kinds of visual information to the operator for SPS type surgical robot. In addition we expect that in the future as we proceed with the development, it will be possible to mount the system on NOTES type robot which has a flexible structure.

References

1. Cavusoglu, M.C., Williams, W., Tendick, F., Sastry, S.S.: Robotics for Telesurgery: Second Generation Berkeley/UCSF Laparoscopic Telesurgical Workstation and Looking Towards the Future Applications. Industrial Robot: Special Issue on Medical Robotics 30(1), 22–29 (2003)
2. Treadt, M.R.: A surgeon's perspective on the difficulties of laparoscopic surgery. Computer-Integrated Surgery: Technology and Clinical Applications, 559–560 (1995)
3. Roviaro, G.C., Varoli, F., Saguatti, L., Vergani, C., Maciocco, M., Scarduelli, A.: Major vascular injuries in laparoscopic surgery. Surg. Endosc. 16, 1192–1196 (2002)
4. Way, L.W., Stewart, L., Gantert, W., Liu, K., Lee, C.M., Whang, K., Hunter, J.G.: Causes and prevention of laparoscopic bile duct injuries: Analysis of 252 cases from a human factors and cognitive psychology perspective. Annals of Surgery 237, 460–469 (2003)
5. Michaelson, M.D., Cotter, S.E., Gargollo, P.C., Zietman, A.L., Dahl, D.M., Smith, M.R.: Management of complications of prostate cancer treatment. CA Cancer J. Clin. 58, 196–213 (2008)
6. Dodgson, N.A., Wiseman, N.E., Lang, S.R., Dunn, D.C., Travis, A.R.L.: Autostereoscopic 3D display in laparoscopic surgery. Computer Assisted Radiology 95, 1139–1144 (1995)
7. Levy, A.: Multi-camera endoscope, US Patent Application Publication 20120053407 A1 (2012)
8. Suzuki, N., Hattori, A., Hayashibe, M., Suzuki, S., Otake, Y.: Development of dynamic spatial video camera (DSVC) for 4D observation, analysis and modeling of human body locomotion. Medicine Meets Virtual Reality 11, 346–348 (2003)
9. Saito, T., Suzuki, N., Hattori, A., Suzuki, S., Hayashibe, M., Otake, Y., et al.: Motion analysis system using DSVC (Dynamic Spatial Video Camera) and 4D human modeling. Computer Assisted Radiology and Surgery, 1376 (2005)
10. Saito, T., Suzuki, N., Hattori, A., Suzuki, S., Hayashibe, M., Otake, Y.: Estimation of Skeletal Movement of Human Locomotion from Body Surface Shapes Using Dynamic Spatial Video Camera (DSVC) and 4D Human Model. Medicine Meets Virtual Reality 14 119, 467–472 (2006)

Hand-Held Force Magnifier for Surgical Instruments: Evolution toward a Clinical Device

Randy Lee[1,4], Bing Wu[2], Roberta Klatzky[2], Vikas Shivaprabhu[1],
John Galeotti[3], Samantha Horvath[3], Mel Siegel[3], Joel S. Schuman[1,5],
Ralph Hollis[3], and George Stetten[1,3,4]

[1] Department of Bioengineering, University of Pittsburgh
[2] Department of Psychology, Carnegie Mellon University
[3] Robotics Institute, Carnegie Mellon University
[4] Department of Biomedical Engineering, Carnegie Mellon University
[5] Department of Ophthalmology, University of Pittsburgh School of Medicine
http://www.vialab.org

Abstract. We have developed a novel and relatively simple method for magnifying forces perceived by an operator using a surgical tool. A sensor measures force between the tip of a tool and its handle, and a proportionally greater force is created by an actuator between the handle and a brace attached to the operator's hand, providing an enhanced perception of forces at the tip of the tool. Magnifying forces in this manner may provide an improved ability to perform delicate surgical procedures. The device is completely hand-held and can thus be easily manipulated to a wide variety of locations and orientations. We have previously developed a prototype capable of amplifying forces only in the push direction, and which had a number of other limiting factors. We now present second-generation and third-generation devices, capable of both push and pull, and describe some of the engineering concerns in their design, as well as our future directions.

Keywords: haptics, touch, robotic surgery, microsurgery, force magnifier, force-reflecting, steady hand.

1   Introduction

A need exists for improvement in the perception of forces by the sense of touch when using tools to perform delicate procedures. One key area for potential applications is ophthalmological surgery, in which we have recently been exploring techniques for image-guided intervention using optical coherence tomography [1]. Therefore, this area in particular has motivated us in the present research. The technology we are developing could, however, also prove helpful in other forms of microsurgery. For example, surgeons routinely repair tiny blood vessels under a microscope that are far too delicate to be felt by the hand of the surgeon. Providing a useful sense of touch for such applications could improve outcome and increase safety in any of these applications. In a broader sense, any sharp tool such as a needle or scalpel is designed

to minimize force between the tool and the tissue, and augmenting the sense of touch in these tools could also prove valuable.

Purely telerobotic systems such as the da Vinci® Surgical System (Intuitive Surgical, Inc.) can provide motion-scaling, so that fine motion of the tool can be controlled by coarser motion of the operator's hand on the controls. Although force at the tool tip cannot be sensed by the operator in the current commercial da Vinci® device, experimental systems have been tested that translate these forces into visual cues [2] as well as into vibrotactile feedback to the operator's fingers [3].

A different, non-telesurgical approach has been demonstrated in several experimental systems, including the Force-Reflecting Motion-Scaling System created by Salcudean, et al. [4] [5], and the Steady Hand Robot described by Taylor, et al. [6][7]. These systems generate a magnified sense of touch by using a robotic arm that holds the surgical tool simultaneously with the surgeon, pushing and pulling as appropriate, to amplify forces detected by small sensors between the handle of the tool and its tip. Because every force requires an opposing force, the robotic arm must be mounted somewhere, and its weight is generally supported by that mounting. Thus in these systems the magnified forces are created between the tool handle and subsequently the floor. To permit free motion of the tool by the surgeon, an elaborate remote-center-of-motion articulated robot arm is required, along with a control system to keep the tool moving naturally, as if controlled just by the operator, so that the surgeon can have something approaching the degrees of freedom and ease of manipulation that he/she is accustomed to with a hand-held tool. Such systems are typically fairly large and complex. Issues arising from the limited and congested workspace typical in microsurgery raise serious challenges to their practical deployment.

The desire to free robotic surgery devices from the floor-standing robotic arm has led to hand-held systems such as the Micron microsurgical instrument from Riviere's group, which uses piezoelectric actuators to move the tip relative to the handle, based on optical tracking of both the tip and handle [8]. The primary goal of Micron is to reduce the effects of hand tremor. It is not suited to provide a magnified sense of touch.

Another hand-held probe is the MicroTactus, developed to enhance tactile sensitivity during minimally invasive surgical tasks such as probing and exploration [9]. The device is instrumented with an accelerometer attached to the tool tip, as well as a solenoid operating inertially against an internal movable weight in the handle to create vibrotactile stimulation. The accelerometer is oriented orthogonal to the actuator to decouple the input signal from the output. Since the forces are generated purely by inertia, they are inherently transitory and must integrate to zero over time. The device is therefore capable of communicating only texture while moving across a surface, rather than non-transitory forces such as those from sustained pushing and pulling against a target.

When the goal is to magnify non-transitory forces for the operator to feel, some external frame to "push against" has generally been required, and researchers have looked for ways of doing that without a table-standing or floor-standing system. The field of haptic simulation faces the same dilemma of generating sustained forces for the fingers to feel without anchoring the renderer to some solid base. Recent examples of such portable solutions include the "active thimble" described by Solazzi, et al. [10].

The device is entirely mounted on one hand. It attaches to the proximal portion of the finger and reaches over to contact the fingertip, thus generating forces between two parts of the operator's own anatomy. As they describe it, "[a] limit of traditional kinesthetic interfaces is the difficulty to achieve a large workspace without a detriment of dynamic performance and transparency or without increasing the mechanical complexity. A possible solution to overcome this problem is to develop portable ungrounded devices that can display forces to the user hands or fingers."

In our approach we extend the concept of ungrounded haptic devices from purely virtual environments to real tools, by including a force sensor for interaction with an actual target. As in the Force-Reflecting and Steady Hand systems described above, we provide a magnified perception via the tool handle of forces sensed at the tool tip. Our approach, however, does not require any freestanding apparatus, and instead produces forces between portions of the operator's own anatomy. The concept is shown in Fig. 1. A hand-held tool contains a sensor, which measures force f between the handle and the tip. This signal is amplified to produce a force F in the same direction on the handle of the tool, using an actuator mounted on the back of the hand (in this case, a solenoid). The human is, in effect, providing a moving platform from which the magnified forces are generated. Since our original publication of this concept [11], a similar approach has been published by Payne, et al., in which forces are generated between the fingertips on an integrated slide in an active tool, relative to the rest of the tool stabilized by other portions of the hand [12].

Fig. 1. The Hand-Held Force Magnifier (HHFM) uses a sensor to measure force f between the handle and the tip, which is amplified to produce a force $F = k f$ in the same direction on the handle using a solenoid mounted on the back of the hand

2   Model-1 Hand Held Force Magnifier

We have previously reported on the Model-1 prototype of the Hand-Held Force Magnifier (HHFM) [11], illustrated in Fig. 2. We briefly review it here. In this initial prototype, the button on a small force sensor (Honeywell FS01, 0-6.7 N) served as the tool tip. The tool handle was the body of a syringe attached to a piece of 1/4 inch brass tubing containing a stack of 8 permanent rare-earth magnets (3/16" Radio Shack 64-1895) inserted into a custom solenoid (75 meters of 30 gauge wire, 25 ohms, approx. 2360 turns). The solenoid was attached by a dual gimbal to a brace, which was mounted to the back of a wrist splint strapped to the operator's right hand. The dual gimbal permitted free rotation in azimuth and altitude, while maintaining a

relatively tight connection for the transmission of force in the range (axial) direction. A control system (not shown) consisted of a linear amplifier capable of supplying 32 V at 2 A, enough to operate the solenoid over its maximum range, to produce a force F from the solenoid of up to 1 N, proportional to, and in the same direction as, the force f sensed at the tool tip (Figs. 1 and 2). Thus we can express the system's behavior as simply

$$F = kf. \tag{1}$$

The proportionality factor k was adjustable from 0 to 5.8. Above this level, the system became unstable.

The Model-1 prototype is shown in Fig. 2 being used to push a spring from the side to bend it. With the gain k set to 0, the spring is felt through the tool to be subjectively quite easy to bend. With k increased to maximum, the spring feels much harder to bend. This haptic illusion is due to the fact that the fingertips must match not only the force of the spring f but also that of the solenoid F. Thus the device magnifies the operator's sensation of touch while keeping the force actually applied to the spring relatively small.
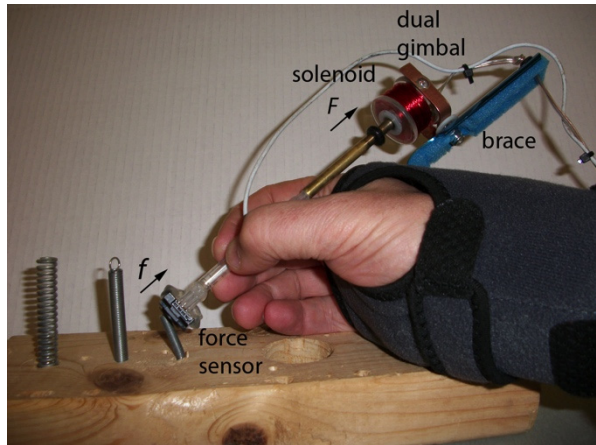
Fig. 2. Model-1 prototype of the Hand-Held Force Magnifier (HHFM). Here the operator "feels" a magnified force generated pushing against a spring.

We hypothesized that the operator can sense forces at the tool tip that are smaller than would otherwise be perceivable, and can control these smaller forces with greater delicacy by interacting with the magnified forces. Two experiments were conducted to characterize the Model-1 HHFM. In the first experiment, the participants' ability to sense the presence of a force (i.e. the absolute detection threshold) was measured with and without the use of the HHFM. In the second experiment, we used a method of magnitude estimation to characterize the impact of the device on the subjective force intensity. Both studies produced a measure of the perceptual magnification of force implemented by the HHFM, and clearly showed that percepts of force at both threshold and supra-threshold levels are rescaled when using the HHFM, demonstrating that the force magnification induced by the device is well perceived by human users [11].

Anecdotally, it was clear that the spring being pushed in Fig. 2 becomes subjectively stiffer when the HHFM is activated. Surprisingly, this is true even when the wrist is not resting on the bench top, at which point the operator is delivering force solely through the shoulder. This was unexpected, since all of the HHFM's magnifying

effect is physically due to forces generated within the hand, and yet the clear perception to the operator is that the muscles of the arm and shoulder are encountering a stiffer spring. Clearly, some form of perceptual integration is occurring.

A number of shortcomings were evident in the Model-1 HHFM. Since the sensor was only able to register positive forces, only pushing interactions could be magnified. The wrist splint used to mount the device on the hand was cumbersome, restrictive, and appropriate only for right-handed people with relatively large hands. The analog circuitry used to control the actuator was limited in terms of its ability to implement more complex control strategies than simple linear gain. Finally, the solenoid tended to overheat and the linear bushing was heavy and prone to binding.

3   Model-2 Hand-Held Force Magnifier

We report here on the recent evolution of the HHFM, as embodied in our Model-2 and Model-3 devices.

In the Model-2 device we replaced the wrist-splint with a simpler and better fitting brace and actuator assembly (see Figs. 3 and 4) constructed of aluminum stock and acrylic tubing. The two-piece hinged brace is secured to the hand by Velcro straps (not shown in Fig. 3), permitting rapid attachment to, and detachment from, the hand. Foam padding on the back of the brace improves fit and comfort. A rotary bearing is press-fit into the top of the brace to allow rotation of the actuator assembly in azimuth. A single hinge allows for rotation in altitude. Instead of a solenoid, we used a commercially available voice coil (LVCM-19-022-02, Moticont, Inc.) as the actuator, capable of generating up to 2.5 N of force in either direction over a 12.7 mm stroke length. In a voice coil, a

Fig. 3. Model-2 brace and actuator assembly

coil of current-carrying wire moves in the magnetic field of permanent magnets inside a stationary housing, as opposed to a solenoid, which has a moving ferromagnetic core acted upon by a stationary coil. Thus a voice coil is capable of quicker response and is less sensitive to the angular dependency on gravity. An aluminum post was secured to the voice coil and supported by a concentric linear bearing to allow the handle to move freely in the axial direction with minimal friction.
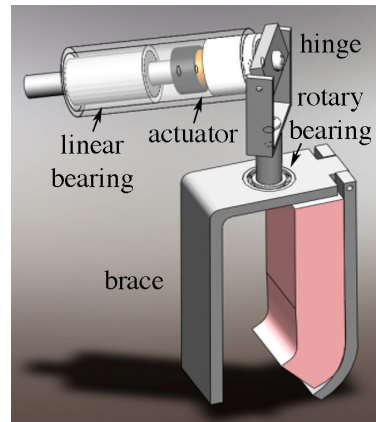
A major goal of the Model-2 was to make it bidirectional, able to amplify both push and pull. Notice that the direction of arrows for both the detected force *f* and the amplified force *F* are reversed in Fig. 4 from Fig. 2, indicating that Fig. 4 depicts an amplified sense of pulling on the test spring. This bidirectional ability was accomplished using the same push-only force sensor used in the Model-1, but with an additional preload spring made of steel shim-stock (see Fig. 4). The spring produces a steady force that can be augmented (push) or relieved (pull) by forces at the tip of

the tool. A small lateral bar was added to the tool tip to permit both pushing and pulling forces to be applied to targets such as the test spring. A pulling force applied to the test spring is shown in Fig. 4.
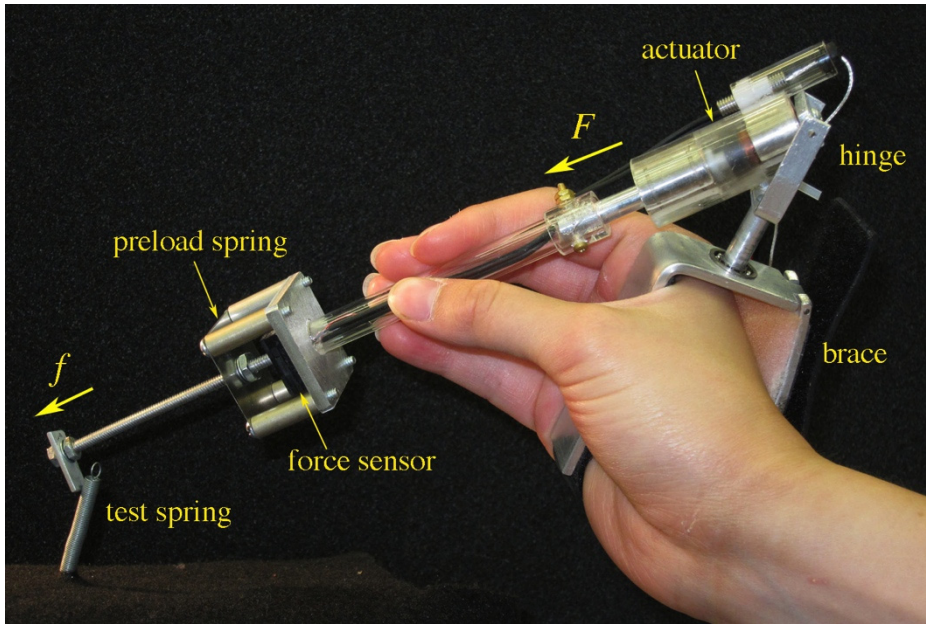
Fig. 4. Model-2 prototype of the Hand-Held Force Magnifier (HHFM), capable of both push and pull. Here the operator "feels" a magnified pull against a test spring.

Another advance in the Model-2 was the incorporation of a microprocessor (ADUC7026, by Analog Devices). The 40 MHz processor has multiple 1 MS/s, 12-bit A/D and D/A convertors. With it, we are able to implement more complex behaviors than with the analog system, using algorithms written in the C programming language, maintaining a 10 kHz through-put without significant jitter. This capability has permitted us to solve the problem of zeroing the preload voltage
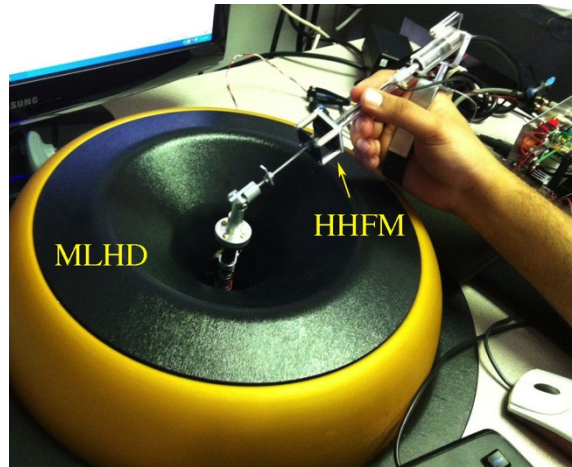
Fig. 5. Magnetic Levitation Haptic Device (MLHD) serving as test bed for the Model-2 Hand-Held Force Magnifier (HHFM)

(analogous to the "tare" button on a typical scale for weighing) by averaging random samples of the sensor voltage over a period of inactivity. The microprocessor furthermore permits rapid exploration of a whole range of more complex functions beyond simple gain, including non-linearity, proportional-integral-differential (PID) control to increase stability, and time varying behaviors that may be required to prevent unwanted tool responses.

We are currently testing the Model-2 HHFM using a Magnetic Levitation Haptic Device (MLHD) (Maglev 200^{TM} from Butterfly Haptics) to accurately control forces and displacements (see Fig 5). The MLHD uses Lorentz forces for actuation, which arise from the electromagnetic interaction between current-carrying coils and magnets. It was developed by co-author Ralph Hollis [13]. Since there are no motors, gears, bearings, or linkages present, the MLHD is free of static friction and able to generate forces precisely with a resolution of 0.02 N. We previously used the MLHD to study simple force magnification with the Model-1 HHFM, and we are now testing the Model-2 using more complex simulations of surgical tasks, such as piercing a sheath of connective tissue or peeling a membrane.

4   Model-3 Hand-Held Force Magnifier

The choice and location of the force sensor is a particularly important aspect in the design of a clinically practical HHFM. The force sensor used in the Model-1 and Model-2 is clearly not ideal, because of its large size. It was chosen for convenience in our proof-of-concept stage, because it comes from the factory pre-calibrated and temperature-compensated. Smaller sensors are available with some extra effort. Also central to the design is the question of optimal sensor location. One possibility is to locate the sensor more proximally, within the handle, and to communicate with the distal tip by means of a mechanical or hydraulic linkage. We have explored the possibility of such a proximal force sensor, using a commercially available pressure
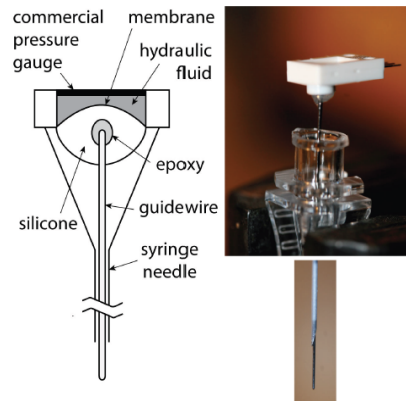
Fig. 6. Guidewire pressure sensor prototype

sensor commonly employed in catheters (Motorola MPX2011DT1). This sensor was chosen for its small size (6.60 mm × 6.07 mm × 3.81 mm), high sensitivity (full scale pressure limit of 75 kPa), and low cost (less than $1). As shown in Fig. 6, the sensor consists of a piezoresistive strain gauge with a pocket of hydraulic fluid held against its forward face by a very thin membrane. The back face of the sensor is exposed to atmospheric pressure, allowing for both positive and negative pressures to be measured relative to the atmosphere. We first explored attaching this sensor to a fluid-filled (water) syringe and transmitting pressure to it via a tiny plunger in the tip of the

syringe needle. This proved ineffective, in that the design of the distal plunger was problematic. We then used a mechanical linkage consisting of a guidewire within the shaft of the syringe, attached to the membrane of the pressure sensor by means of a small epoxy droplet within a flexible silicone mantle (again see Fig. 6). The pressure sensor requires pre-amplification, which we accomplished using an operational amplifier mounted just behind the sensor in the handle, for optimal signal-to-noise. The entire apparatus, attached to the Model-2 brace and actuator assembly, is shown in Fig. 7. This overall design proved quite sensitive to both pushing and pulling, but suffered from significant hysteresis due to friction of the guidewire against the inside of the needle.
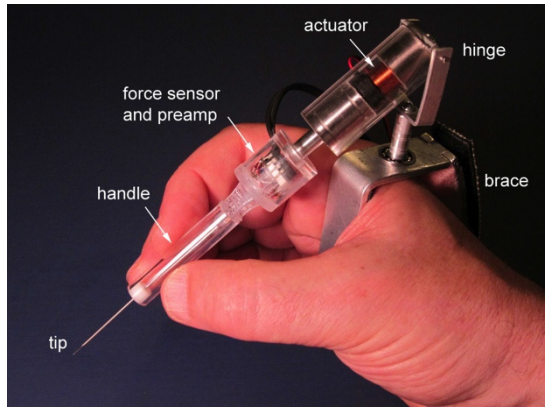
Fig. 7. HHFM design with proximal pressure sensor and guidewire linkage

To address this issue we adopted a different approach to incorporating the pressure sensor, which maintained the mechanical linkage while eliminating the problems with friction. In what we now call the Model-3 Hand Held Force magnifier (see Fig. 8), force is transmitted from the tip to the pressure sensor via a mechanical linkage consisting of a plastic rod. The rod is stabilized by what is sometimes known as a "spider," shown in detail in Fig. 8 A and B, which permits axial motion without permitting lateral motion. The rod becomes narrower as it enters a hole surrounding the force sensor, and the space within is filled with silicone. Unlike the system shown
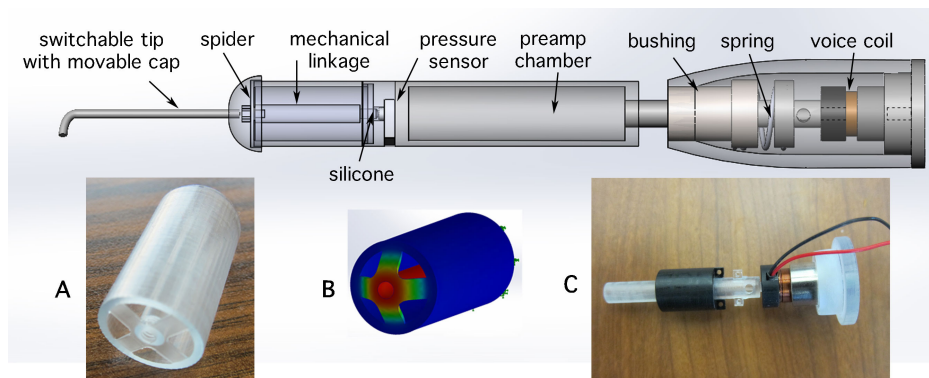
Fig. 8. Model-3 HHFM with removable tip. Spider is shown from front in **A** and displaced under simulated force in **B**. Force is transmitted to pressure sensor via a mechanical linkages and silicone deposited on face of sensor. Spring keeps voice coil at rest in the center of its range. Voice coil, bushing and mounting holes for spring are shown in **C**.

in Fig. 6, no epoxy bead is required, as both pushing and pulling (suction) are transmitted to the pressure sensor by the silicone in the small chamber. The combination of the spider and the rim of the small chamber in front of the pressure sensor reduces the effect of lateral forces at the tip on the sensor and provides a sturdy housing to protect the delicate sensor. Construction by stereolithography (SLA), permits major sections (such as the housing, spider, and rod in Fig. 8 A) to consist of a single piece of plastic.

Our initial tests of the sensor linkage are promising. Superimposed on a self-noise for the sensor of approximately 0.2 mN, we have found temperature-related drift to be less than 5 mN/minute in a common room-temperature environment, with less than 5 mN of hysteresis in the sensor when the tool tip is cyclically loaded and unloaded. This is a significant improvement (approximately 10-fold) over the hysteresis we experienced due to friction in the guidewire system. Since we can continually recalibrate (tare) the sensor to zero during brief periods of inactivity, such temperature drift and hysteresis are not expected to be a significant problem, and thus we can reliably amplify detected forces in the 1 mN range. Furthermore, the design has proven extremely robust, without any damage experienced thus far to the pressure sensor through normal handling of the tool, although we have not yet conducted destructive testing. Off-axis forces are attenuated by a factor of at least 20.

An ongoing concern in previous models involved the actuator. Voice coils, as well as conventional solenoids, produce forces that vary with translation of the coil relative to the ferromagnetic element. For this reason, producing known forces requires either measuring that translation or limiting it to a very small range. Furthermore, the range of motion for the voice-coil is limited in both the pushing and pulling directions, and once either limit is reached, force magnification in that direction by the HHFM is no longer possible.

We have developed a method of addressing these issues in the Model-3. The voice coil actuator is now stabilized by a simple expansion/extension spring mechanism, which solves the two issues of knowing where the coil is relative to the magnet and not permitting the coil to reach either hard limit in its range of motion (see Fig. 8, top and **C**). No matter where the operator's fingers grip the handle along its length, the operator naturally relaxes against any constant displacement of the spring in either direction from its central resting state. In that state, the spring exerts no force and thus does not diminish the force that can be exerted by the voice coil, provided that such force is matched by an equal and opposite force from the fingers. This seems to be what happens during use of the HHFM, although we have not yet made a detailed measurement of the actual displacement. In any event, we have not had any further instances of the voice coil reaching the limits of its range, and the operator no longer has to be concerned with keeping it in the center of its range, which had been a significant distraction in previous models.

Figure 9 shows the functional Model-3 HHFM in the hand of an operator. At present, we are still using the same brace as in the Model-2 device, though we are planning to streamline this portion of the device to a much simpler attachment to a surgical glove. The Model-3 is electronically interchangeable with the Model-2, and currently uses the same microprocessor-based control circuitry.
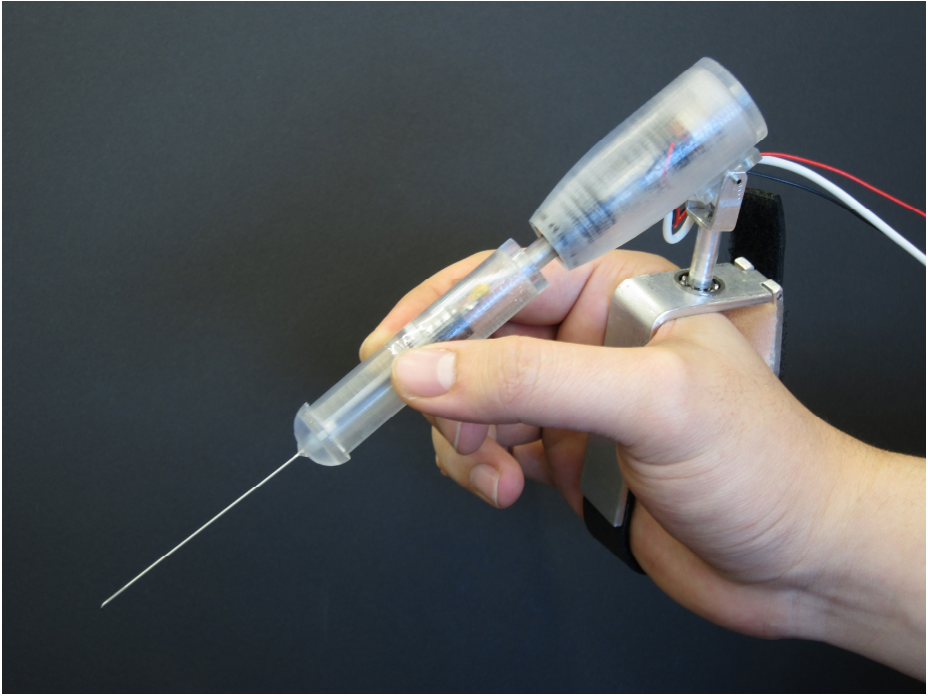
Fig. 9. Model-3 HHFM shown in operator's hand. The same brace as in the Model-2 device is still present, though a surgical glove attachment is under development.

5   Discussion

When we initially developed the HHFM, we were considering the clinical need for greater sensitivity to forces during eye surgery, but many applications may actually be suited for this technology. We have listed some of these in Table 1. Most involve rigid tools, such as needles, scrapers, hooks, scalpels and blunt dissectors used in microsurgery, where forces may be so delicate as to be impossible to feel. However, we have come to understand that such minute forces may be present in regular surgery as well, especially with sharp tools designed, after all, to minimize the forces resisting cutting and stabbing. Furthermore, potential uses of the HHFM may include operating in the "bloody field" of cardiac surgery, where a magnified sense of touch may improve the surgeon's ability to feel structures when vision is obscured by blood. The technology may also be adapted for use at the end of a catheter, permitting axial force and axial torque to be felt from the handle of the device, for example, to navigate the branch points in a vein or bronchus.

The question of whether to amplify non-axial forces and torques (those not along the axis of the tool) also arises. Certainly, lateral forces at the tip of a rigid tool do exist and are largely converted to torques at the handle. Thus there are effectively, two, instead of four, remaining degrees of freedom: force/torque in azimuth and

force/torque in altitude. The resulting mechanical disadvantage of effectively operating from the short end of a lever makes amplification of such forces/torques inefficient. Another way to look at this is that non-axial forces/torques are inherently already magnified, in the sense that they are counteracted only with significant mechanical disadvantage.

Another issue that has been raised is the constraint imposed upon the operator by the attachment to the hand, such that the operator must grip the handle of the tool at a particular location to be comfortable, and thus the tip cannot be moved freely in the axial direction. While this is true, most tools already have a preferred location for holding, and the operator generally moves the entire hand to effect axial motion of the tool tip.

Sterility will be a concern in most surgical applications. We envision building the brace and actuator into a surgical glove in such a way that various portions may be detached, some being reusable and

Table 1.

Eye Surgery
• Cataract surgery
• Vitrectomy
• Surgery for Glaucoma
• Retinal surgery
• Corneal surgery
• Retrobulbar Injection
• Intra-ocular tumor excision
Other Microsurgery
• Neurosurgery
• Vascular & Microvascular Surgery
• Plastic Surgery
• Trauma and Reconstructive Surgery
Other Surgical Applications
• Epidural needle insertion
• Jugular vein needle insertion
• Cardiac surgery (bloody field)
• Minimally Invasive Surgery (laparoscopic, cardiac)
• Catheter based vascular procedures

sterilizable, while others are pre-sterilized and disposable. In particular, removable disposable tips would be a practical solution to the need for a variety of force-magnified tools during a given procedure.

Finally, in the Model-3 HHFM we moved the sensor into the handle to reduce the bulkiness of the distal portions of the tool. Another approach we are considering for future prototypes is to keep the force sensor distal to the handle but to greatly reduce its size. Very small piezoresistive surface-mount pressure sensors are available (for example, 2.5 mm x 3.3 mm x 1.3 mm, 3000 series from Merit), which could be enclosed in the distal portion of a catheter-based system. Moving sensors closer to the tip reduces interactions between orthogonal forces and torques. These may also be disambiguated by micro-machined arrays of strain gauges, such as those developed by Berkelman, et al. [14]. For distal sensing, another appealing solution is the optical Bragg sensor, which is small enough to be built into a fine needle tip, and which may be interrogated via optical fiber. In one such system being developed by Sun, et al., for use in retinal surgery, mounting the sensor in the tip eliminates confounding forces resulting from insertion through the sclera [15]. This would be true in our proposed catheter-based system as well.

6   Conclusion

We have reported here on progress with the HHFM, including the design and construction of several new working prototypes, as the technology evolves towards a practical clinical device. We have also discussed some likely applications in clinical medicine. The major contribution of our work, we believe, is to provide a magnified sense of touch without requiring an external robotic arm. The force that was generated between the operator's hand and the floor by the robotic arm in previous implementations of force magnification is replaced by a force generated between two locations on the operator's hand, freeing the design to permit a small, light, hand-held device.

References

1. Galeotti, J., Sajjad, A., Wang, B., Kagemann, L., Shukla, G., Siegel, M., Wu, B., Klatzky, R., Wollstein, G., Schuman, J., Stetten, G.: The OCT penlight: In-situ image guidance for microsurgery. SPIE Medical Imaging, paper #7625-1 (2010)
2. Bethea, B., Okamura, A., Kitagawa, M., Fitton, T., Cattaneo, S., Gott, V., Baumgartner, W., Yuy, D.: Application of Haptic Feedback to Robotic Surgery. J. Laparoendosc. Adv. Surg. Tech. A 14(3), 191–195 (2004)
3. Kuchenbecker, K.J., Gewirtz, J., McMahan, W., Standish, D., Martin, P., Bohren, J., Mendoza, P.J., Lee, D.I.: VerroTouch: High-Frequency Acceleration Feedback for Telerobotic Surgery. In: Kappers, A.M.L., van Erp, J.B.F., Bergmann Tiest, W.M., van der Helm, F.C.T. (eds.) EuroHaptics 2010, Part I. LNCS, vol. 6191, pp. 189–196. Springer, Heidelberg (2010)
4. Salcudean, S.E., Yan, J.: Motion scaling teleoperating system with force feedback suitable for microsurgery, U.S. Patent 5,382,885 (1995)
5. Salcudean, S.E., Yan, J.: Towards a Force-Reflecting Motion-Scaling System for Microsurgery. In: IEEE International Conference on Robotics and Automation, San Diego, California (1994)
6. Taylor, R., Jensen, P., Whitcomb, L., Barnes, A.C., Kumar, R., Stoianovici, D., Gupta, P., Wang, Z., deJuan, E., Kavoussi, L.R.: A Steady-Hand Robotic System for Microsurgical Augmentation. In: Taylor, C., Colchester, A. (eds.) MICCAI 1999. LNCS, vol. 1679, pp. 1031–1041. Springer, Heidelberg (1999)
7. Fleming, I., Balicki, M., Koo, J., Iordachita, I., Mitchell, B., Handa, J., Hager, G., Taylor, R.: Cooperative Robot Assistant for Retinal Microsurgery. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 543–550. Springer, Heidelberg (2008)
8. Tabars, J., MacLachlan, R., Ettensohn, C., Riviere, C.: Cell Micromanipulation with an Active Handheld Micromanipulator. In: 32nd Annual International Conference of the IEEE EMBS, Buenos Aires, Argentina (2010)

 9. Yao, H.-Y., Hayward, V., Ellis, R.E.: A Tactile Enhancement Instrument for Minimally Invasive Surgery. Computer Aided Surgery 10(4), 233–239 (2004)
10. Solazzi, M., Frisoli, A., Bergamasco, M.: Design of a Novel Finger Haptic Interface for Contact and Orientation Display. In: IEEE Haptics Symposium, Waltham, Massachusetts, March 25-26, p. 129 (2010)
11. Stetten, G., Wu, B., Klatzky, R., Galeotti, J., Siegel, M., Lee, R., Mah, F., Eller, A., Schuman, J., Hollis, R.: Hand-Held Force Magnifier for Surgical Instruments. In: Taylor, R.H., Yang, G.-Z. (eds.) IPCAI 2011. LNCS, vol. 6689, pp. 90–100. Springer, Heidelberg (2011)
12. Payne, C., Latt, W.: A New Hand-Held Force-Amplifying Device for Micromanipulation. In: 2012 IEEE International Conference on Robotics and Automation, Saint Paul, Minnesota, May 14-18 (2012)
13. Hollis, R.L., Salcudean, S.E.: Lorentz Levitation Technology: A New Approach to Fine Motion Robotics, Teleoperation, Haptic Interfaces, and Vibration Isolation. In: 5th International Symposium on Robotics Research, Hidden Valley, PA, October 1-4 (1993)
14. Berkelman, P.J., Whitcomb, L.L., Taylor, R.H., Jensen, P.: A Miniature Microsurgical Instrument Tip Force Sensor for Enhanced Force Feedback during Robot-Assisted Manipulation. IEEE Transactions on Robotics and Automation 19(5), 917–922 (2003)
15. Sun, Z., Balicki, M., Kang, J., Handa, J., Gehlbach, P., Taylor, R., Iordachita, I.: A Sub-Millemetric, 0.25mN Resolution Fully Integrated Fiber-Optic Force Sensing Tool for Retinal Microsurgery. Int. J. Comput. Assist. Radiol. Surg. 4(4), 383–390 (2009)

A Simulator for Designing Control Schemes for a Teleoperated Flexible Robotic System

Antonio De Donno*, Florent Nageotte, Philippe Zanne, Laurent Goffin, and Michel de Mathelin

Equipe AVR, LSIIT UMR 7005, University of Strasbourg and CNRS,
1, Place de l'Hôpital, 67091 Strasbourg cedex, France
`adedonno@unistra.fr`

Abstract. No-scar surgery, which aims at performing surgical operations without visible scars, is the vanguard in the field of Minimally Invasive Surgery. No-scar surgery can be performed with flexible instruments, carried by a guide under the vision of an endoscopic camera. This technique brings many benefits for the patient, but also introduces several difficulties for the surgeon. We aim at developing a teleoperated robotic system for assisting surgeons in this kind of operations. In this paper we present this novel system, its kinematics and the strategies that we are proposing for its control. Also, we present a virtual simulator that we have specifically developed for our system with the purpose of assessing the control strategies and studying the possible system mechanical issues. Trials with novices show that the best control strategies depend on the kind of task to be performed and on the presence of simulated errors.

Keywords: Virtual simulation, surgical robotics, No-scar surgery, telemanipulation.

1   Introduction

Nowadays, the vanguard in the minimally invasive surgery field is represented by the 'No-Scar Surgery', that encompasses all the techniques that permit to perform a surgical operation without visible scars. In the field of no-scar surgery there are two common techniques:

- Natural Orifice Transluminal Endoscopic Surgery (NOTES) [1]: special technique for abdominal surgery in which instruments are passed via natural orifices (mouth, anus, vagina).
- Single Port Access (SPA) Surgery [2]: in this technique the surgeon makes only one skin incision, typically inside the umbilicus (for abdominal surgery) or below the armpit (for thoracic surgery). Compared to NOTES, the instruments are shorter, since they are introduced near the operation field, and they can be rigid or flexible.

---

* Corresponding author.

Compared to traditional laparoscopy, these two techniques represent an impressive improvement for the patient, because they increase the positive effects of traditional laparoscopy and, in addition, they provide better aesthetic results.

Unfortunately, beside the advantages, no-scar surgery greatly increases the difficulties for the surgeon: the space for the instruments is smaller than in other techniques (triangulation is one of the main positioning problems), a long training period is necessary to regain hand-eye coordination (in many cases surgeons conduct operations with the hands crossed [5]) and the operation mean time is longer. Furthermore, this approach requires at least two surgeons to manipulate the instruments.

In this scenario, robotics could assist surgeons during operations, improving control, precision, ergonomics and allowing only one surgeon to telemanipulate the whole system. We are currently developing a novel robotized flexible system suitable for no-scar surgery. We have also developed a virtual simulator that replicates the kinematic of our system, with multiple objectives. First, the simulator can allow to test different control strategies without the need to deal with known practical problems such as mechanical non-linearities or sensing limitations. This can provide us with the best control strategies and help us focus on solving the underlying physical limitations. Second, the simulator can be used to gradually introduce imperfections and monitor their effect on the user capabilities. Finally, the simulator can be used by surgeons to familiarize with the control of the robotic system, with the possibility to evaluate their skills and to choose their preferred control modes.

In this paper we first briefly present in section 2 the manual system that we have robotized. Then, in section 3 we describe our teleoperated system in all its parts, focusing on kinematics, control strategies, master input and mechanical non-linearities that can be assessed using the simulator. Finally, the virtual simulator and its features are presented in section 4, together with a discussion about the first results of two path following tasks performed with it.

2   System Description

Our prototype is based on the Anubiscope system, manufactured by the Karl Storz Endoskope GmbH company. Anubiscope (see fig. 1) is a flexible multipurpose platform for endoluminal and transluminal procedures that permits to introduce up to three flexible instruments inside the patient from a single entry point.

This flexible guide consists of three major parts: an handle, where control knobs and instruments channels entrances are located, a long flexible part and a controllable tip, that integrates the endoscopic camera and a head with an articulated insertion system (see fig. 2). Thanks to its controllable flexible tip, the endoscope can be bent in two perpendicular directions (Degrees Of Freedom (DOFs) noted 2, 3, see fig. 3). The endoscope movements are assured by knobs, located at the distal part of the flexible guide. Each knob is attached to a couple

Fig. 1. Karl Storz Anubiscope system

of antagonistic cables: when the knob is rotated in one direction, one cable is pulled, which produces the bending movement. The particular head design permits to realize triangulation, needed to correctly position the instruments during the operation. This system is used with flexible instruments, which can be bent in one direction and which can be rotated and translated inside their channels (DOFs noted 4, 5, 6, see fig. 3). Typically, this system requires at least two surgeons to operate it: the main surgeon controls the instruments by their handles, while the second surgeon moves the endoscope according to the main surgeon requests. This operation implies a good coordination between surgeons.

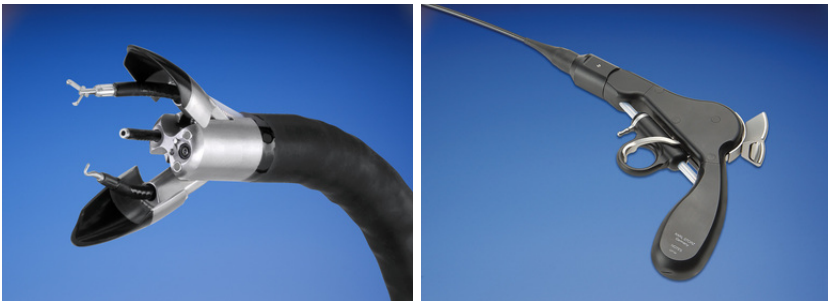**Fig. 2.** Anubiscope head (left) and instrument handle (right)

3   Teleoperated System

Our project aims at developing a teleoperated system for no-scar surgery based on the Anubiscope platform described in section 2. In the teleoperation scheme,

the robotized Anubiscope will become the slave part, controlled by two master interfaces held by the surgeon. There are several advantages in the use of a teleoperation approach:

- with an appropriate master system, it is possible to leave the complete control of the slave system to a single user;
- the slave system behavior could be changed according to the chosen control strategy;
- the user input could be filtered and treated before being sent to the slave system, in order to reject hands tremor or to scale the movements.

But the teleoperation approach also introduces some difficulties. Especially, since master and slave systems usually have different kinematics, it is necessary to find a good mapping between them to allow good control accuracy and ease of use.

3.1 Slave System

The Anubiscope robotization consisted in the substitution of all the manual DOFs with electric motors, chosen depending on velocities and torques specifications and numerically controlled (see for instance [9] for a very first version of a close system). Globally, our system will have the following DOFs (see fig. 3):

- 3 DOFs for the endoscope: one translation (noted 1, Z direction) and two deflections (noted 2 and 3, X and Y directions);
- 3 DOFs for each instrument: one translation (noted 4), one rotation (noted 5) and one deflection (noted 6);
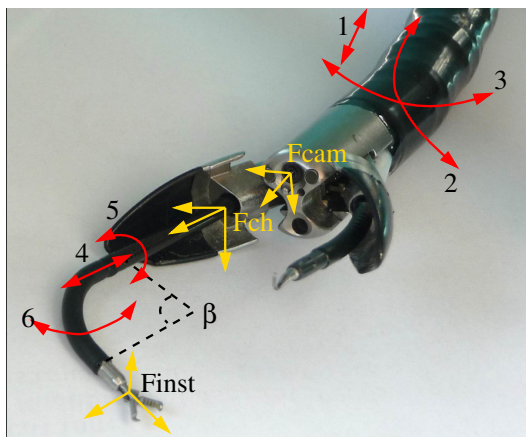- another additional DOF for actuated instruments (like graspers).

Fig. 3. System DOFs - 1: Endoscope translation, 2,3: Endoscope deflections; 4: Instrument translation; 5: Instrument rotation; 6: Instrument deflection

Each deflection movement is possible thanks to a vertebrae chain controlled by a couple of parallel antagonistic cables. The cables, that pass inside the chain, are attached to the distal vertebra. Each following vertebra is passively connected to the previous one and can rotate in one direction. To bend a flexible section of the system (either the ones in the endoscope or in the instruments) it is necessary to pull one cable and to keep the other unstrained: this will apply a torsion moment to the distal vertebra that will be mechanically transmitted to the following ones.

Kinematics. The relationship between Cartesian and joints spaces for the slave system can be obtained through the Jacobian or with the position Inverse Kinematics (IK):

$$\dot{q} = J^{-1}\Omega \qquad q = f^{-1}(X)$$

where q is the m-vector of joint positions and $\dot{q}$ its time derivative, X the n-vector of Cartesian positions and Ω the corresponding Cartesian speeds, J the (n x m) Jacobian and $f{:}\mathbb{R}^m \to \mathbb{R}^n$ the position Forward Kinematics (FK).

To compute these formulations it is necessary to describe the kinematic structure of the robot, by means of rigid transformations between fixed frames. Although flexible bending sections cannot be directly represented by the usual Denavit-Hartenberg (DH) convention, it has been shown (Hannan *et al.* [6], Jones *et al.* [7]), that it is possible to provide a DH parametrization to these "continuum" systems using some simplifications: assuming a constant curvature and an inextensible length of the formed arc, each bending section can be modeled by 5 interdependent virtual joints, 4 rotations and 1 translation. In this way, one can obtain a transformation matrix for every bending part and calculate the global FK transformation as in the conventional manner.

From the FK model, the Jacobian matrix can be easily computed analytically or numerically. On the contrary, for our system there is no closed form solution for the position IK. We propose an approximate solution to this problem using the analysis of the instrument workspace (WS). When the instrument base position is fixed, the remaining DOFs (deflection and rotation) allow the tip of the instrument to span a surface in the Cartesian space, that can be approximated by an ellipsoid (shown in red on fig. 4). This local instrument WS, centered at (0, 0, 7.359) mm with respect to the instrument flexible section base, is truncated because of the deflection limit. When using the translation, the ellipsoid is rigidly translated and the global instrument WS is described by a cylinder with ellipsoidal caps (shown in blue on fig. 4).

Given a desired position (X, Y, Z) of the tip of one instrument expressed in the Cartesian space, our IK algorithm checks if the desired position (X', Y', Z'), expressed in the frame F_{ch} attached to the instrument channel exit (see fig. 3), lies inside the global WS. If it is outside, a deflection of the endoscope is computed to bring the point inside the instrument WS. Otherwise, the translation of the instrument is obtained by computing from (X', Y', Z') the center C of the ellipsoid corresponding to the local WS. The instrument rotation can be easily
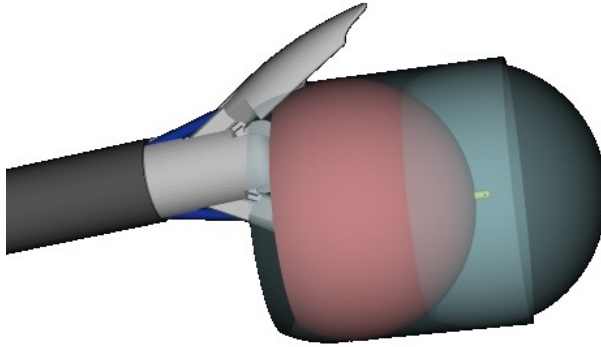
Fig. 4. Instrument workspace: in red the workspace generated by deflection and rotation only, and in blue the complete workspace with translation
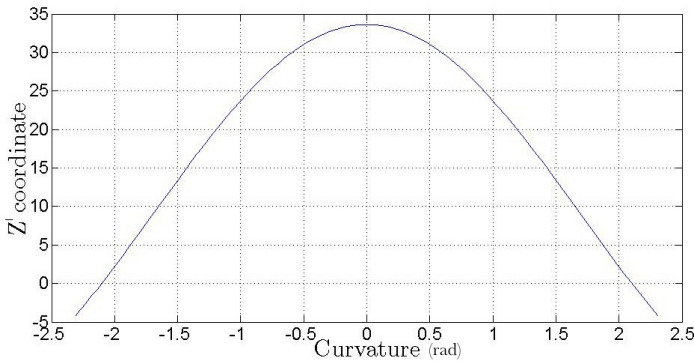
Fig. 5. Instrument local Z' coordinate in function of instrument curvature angle

computed with: $\theta = atan2(Y', X')$. Finally, the link between the local position Z' of the instrument tip and the angle of curvature of the instrument (as shown on fig. 5) is used to obtain the instrument deflection.

Mechanical Non-linearities. Due to its design, the Anubiscope manual system is affected by one major mechanical non-linearity, namely backlash in the deflection actuation systems. This non-linearity, that also appear in the robotized system, arises from the deflection actuation. As explained in section 3.1, the bending part of an instrument is controlled by a couple of antagonistic cables about 1 m long. To deflect the instrument in its bending plane it is necessary to pull one cable and to keep the other unstrained. But if one wants to change the bending direction, it is necessary first to put in tension the unstrained cable before obtaining a movement. In the meanwhile, the measured motor position

changes without apparent movements of the instrument distal tip. This non-linearity is complex and depends on several parameters (see [10]) and hence cannot be easily included in the kinematic models presented previously.

3.2 Slave Control Strategies

In this section we will present the strategies that we are proposing to control one single instrument. Our goal is to offer a Cartesian control of the instrument tip because it is more intuitive and natural for the user than a joint per joint control. Also, the control will be made in the frame F_{cam} attached to the camera (see fig. 3), because the image is the only feedback available for the user. All the control strategies only rely on the kinematic models and do not take into account backlash non-linearities.

Open Loop, Speed Control (see fig. 6): the desired instrument tip Cartesian speed Ω is directly multiplied by the inverse of the Jacobian J^{-1}, in order to obtain the corresponding speed references $\dot{q}$ for the motors.
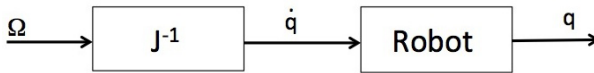
Fig. 6. Open loop, Speed control

Closed Loop, Cartesian Control (see fig. 7): the desired instrument tip Cartesian position X_{ref} is compared with an estimated robot Cartesian position $\tilde{X}$, and a proportional control law λ with an additional feedforward term is used to compute the Cartesian velocity input, which is then fed to the inverse of the Jacobian J^{-1}. Practically, the estimate of the actual Cartesian position $\tilde{X}$ can be obtained by an external sensor (such as a magnetic tracking system).
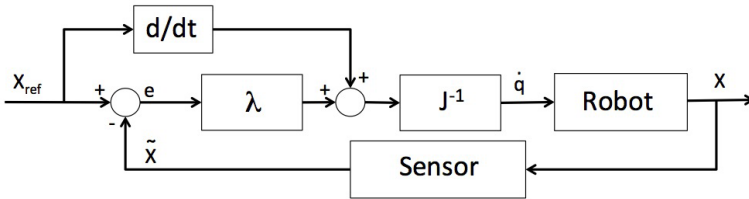
Fig. 7. Closed loop, Cartesian control with feedforward

Closed Loop, Joint Control (See Fig. 8): in this case the loop is closed at the joint level; to calculate the reference joints position q_{ref} from the desired Cartesian position X_{ref} we use the IK algorithm described in section 3.1. This algorithm provides a set (minimum two, if the point is reachable by the instrument tip) of joints configurations for a desired Cartesian position. The reference joints position q_{ref} is then compared with the actual one q, provided by the motor encoders, to calculate an error and to use it as input for the closed loop scheme.

Fig. 8. Closed loop, Joint control

3.3   Master System

The master system is composed by two Force Dimension Omega7 haptic interfaces. Each interface offers 7 DOFs: 3 active translations, 3 passive rotations and 1 active trigger. Since each instrument has only 3 DOFs, only the master interface position is used for controlling them. The user input can be obtained through the interface positions, expressed in the i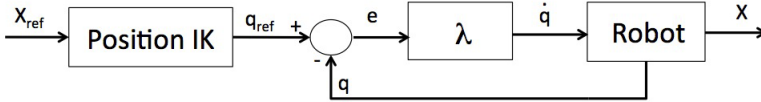nterface frame, or through the velocities applied on them. Since the slave control schemes can require either a position reference or a speed reference, four different situations must be considered:

- master interface position, position reference for the slave system: the master interface workspace is properly scaled to fit the instrument workspace; in this manner, the position of the master interface can be considered as an absolute position reference for the slave system;
- master interface position, velocity reference for the slave system: the master interface position is intended as the speed reference for the system; the interface rest position is equivalent to null speed;
- master interface velocity, position reference for the slave system: in this case, an integration of the master interface velocity is computed to get the desired position;
- master interface velocity, velocity reference for the slave system: the user movements on the master interface determine the speed applied to the slave system. In this case, when the master interface is stopped the corresponding velocity input is null.

The user input is obtained by means of a software library that allows to communicate with the interfaces and to access to their encoders positions. Before sending the user input to the control schemes we apply a scaling filter, to reduce or magnify user movements, and a movement filter, to reject the natural hands

tremor and the noise on the interface encoders. The movement filter acts as a dead zone in the neighborhood of the interface rest position.

According to the slave scheme input and the master interface input types (position or velocity), we also reproduce a force feedback effect on the haptic interfaces:

- Spring: the force feedback acts as a spring that, at any times, drives the haptic interface toward its rest position, in order to send a null speed reference when the user leaves the interfaces.
- Error: the force applied on the interface is proportional to the error vector between the instrument position and the interface position; this force feedback effect should avoid large interface movements that could not be followed by the instruments because of the implemented motors speed limitations.
- Gravity compensation: a zero resultant force is applied on the interface, in order to compensate its weight.

On table 1 a summary of all the possible combinations of slave inputs, master inputs and applied force feedbacks is presented.

Table 1. Force Feedback effects

Slave Input	Master Interface Input	Force Feedback
Speed Ω	Position	Spring
	Speed	Gravity compensation
Position X_{ref}	Position	Error
	Integrated speed	Gravity compensation

4   Simulation

All the control strategies described in section 3.2 have been implemented in a virtual simulator that we have specifically developed to better understand the system behavior. However, the simulator is not aimed at being a perfect model of the robot: it is possible to make the simulation realistic introducing errors like encoder offsets, model errors, frictions or delays in the estimate. In this way, we can also evaluate the user responses in both ideal and real cases. To assess the different proposed strategies some test environments have been created.

4.1   Simulator Features

The simulator is completely written in C#, while the graphic engine is provided by the VtkDotNet libraries. At the startup, two main windows are presented to the user: the main form, that contains all the commands to load specific simulator features, and the graphical 3D window. There are two possible graphical views: an external view, in which the user can see the system virtual model, and a camera view, that replicates the surgeon view during an operation (see fig. 9).
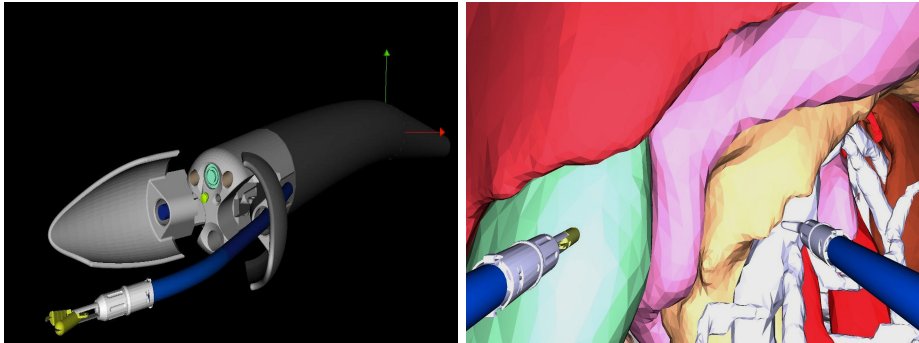
Fig. 9. External view (left) and Camera view (right) on the virtual simulator

The user interaction can be performed with the haptic interfaces, presented in section 3.3, connected via USB ports. With this simulator the user can freely control the whole virtual system with the master interfaces, to become familiar with it, and he/she can also execute some common tasks, like follow a path or simulate a suture movement. Each simulation is stored on the hard disk, so that at any time the user can review his/her actions and the temporal evolution of all the most important system parameters (see fig. 10).
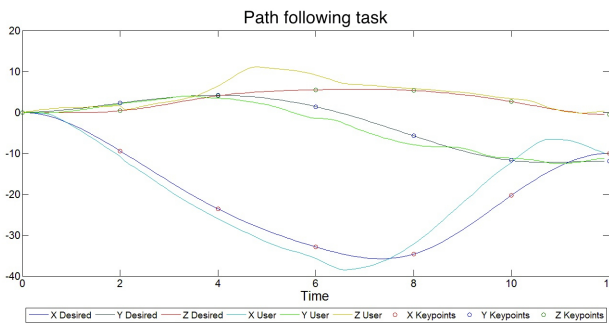
Fig. 10. Results after a path following task. Curves show the desired and actual paths along x,y, and z axes. Circles indicate keypoints.

The main endoscope and the two instruments are implemented inside the simulator by means of two classes each: the first class contains the kinematic model of the system part, while the second class manages the graphical model. In this way, the kinematic computations can run at an higher frequency than the graphical refresh, that normally requires high workloads. An important role in the simulation is covered by the class that simulates the motors behavior. It is possible to use a perfect model of the motor, that reaches immediately the

requested speed without need to accelerate or decelerate, or a realistic model with a trapezoidal speed profile (linear accelerations and decelerations with different values). In the motor class we have also implemented encoder offsets, which represent uncertainties on the relative position calibration between instruments and endoscope, and the backlash, that appears during a motor rotation direction change as a motor movement that produces no graphical effect on a given range.

4.2   Path Following Task

With our simulator we want to assess the slave system behavior according to the control strategy applied and the master interface input. Particularly, we are interested to evaluate the accuracy of the gesture and the response speed of one single arm in both ideal and real cases, in order to understand how much the non-linearities influence the system control. A good way to do this could be to define a path that the user has to follow with one instrument. In our simulator, a path is defined as the cubic interpolation of a number of key points placed in the virtual space. It is possible to load a previously stored path, or to define a brand new one by directly clicking with the mouse on the virtual scene and putting some key points. To obtain a realistic effect, the key points are attached over a 3D model of a human organ, in order to obtain a path drawn over the organ surface (see fig. 11). This helps also to improve the depth perception of the virtual scene.

Fig. 11. Precision task (left) and Simple task (right)

During the task, the instrument position is sampled at a constant rate and stored in memory, as well as the execution time is recorded. When the task is accomplished, an offline routine performs a warping, based on the key points, between the user and the interpolated paths, thus providing an error vector. So we can define a score function as a weighting of the execution time (weight ϕ_T) and the error between the desired and the user paths (weight ϕ_e):

$$Score = \phi_T T + \phi_e \sqrt{e_X^2 + e_Y^2 + e_Z^2}$$

This score function allows to evaluate and compare different gestures according to the type of task executed. In order to evaluate accuracy and response speed, two tasks have been implemented inside the simulator (see fig. 11):

- **Precision task**: the path has 4 key points and mimics a motion during a suturing task. The precision in the gesture represents the major part of the final score ($\phi_T = 1$, $\phi_e = 10$).
- **Simple task**: the path is a large simple motion. An error, shown as a cylinder around the desired path, is tolerated, and the aim is to accomplish the task as fast as possible ($\phi_T = 10$, $\phi_e = 5$).

4.3   Task Protocol and Results

We have conducted a preliminary test phase involving two groups of ten persons each with no previous experience on the simulator. The first group tested the encoder offset error for the instrument rotation, while the second group tested the backlash non-linearity in the instrument deflection. We have chosen the values of 20 degrees for the rotation encoder offset and 2 mm of cable attached to the motor at the proximal side for the deflection backlash: both values come from a direct evaluation on the real system.

Each user starts with an initial apprenticeship phase, in order to familiarize with the control laws and the force feedback applied on the haptic interface. In this phase, that lasts three minutes, the user is guided to perform simple movements and to understand the differences among the proposed control laws. After this initial phase, the Precision task is proposed: the user has to complete it with the different combinations of slave control laws and master interface inputs chosen at random, first without simulated non-linearities and then adding the error. When the Precision task is completed, the Simple task is proposed with the same rules.

The final results of this test session are summarized in Table 2. Each task must be evaluated separately, because the scores are computed with different weights. The two groups obtained similar scores in ideal conditions, so we reported a mean global score for both. This denotes the same starting level of both groups and allows us to make a direct comparison of errors effects in a specific task.

In ideal conditions, a speed control strategy with a position master input permits to obtain better results for the Precision task. After an initial adaptation to the control method, the most common user feedback is that this modality permits to perform precise gestures, because the user input is given as small Cartesian velocity variations with respect to the actual instrument Cartesian position. The counterpart is a lower instrument reactivity for large requested movements, and this is the reason why a position control strategy provides better results in the Simple task.

The observations change significantly when an error is introduced inside the simulation. In Table 3 we reported the results of the second group with the strategies that involved a position master input.

Table 2. Path following results for both considered tasks

Precision task				
Slave Strategy	Master Input	Mean Score (no error)	Mean Score (rot error)	Mean Score (backlash)
OL Speed	Position	75.268	143.265	151.029
	Speed	77.056	117.600	507.05
CL Joint	Position	87.317	89.141	184.342
CL Cartesian	Position	83.345	104.262	110.255
	Int. Speed	84.110	88.683	199.71

Simple task				
Slave Strategy	Master Input	Mean Score (no error)	Mean Score (rot error)	Mean Score (backlash)
OL Speed	Position	80.451	81.726	92.577
	Speed	91.879	149.493	101.64
CL Joint	Position	77.131	92.958	153.863
CL Cartesian	Position	72.268	99.290	88.591
	Int. Speed	78.754	129.780	143.820

Table 3. Comparison between the results of the second group (the error added is the backlash)

	Precision task			Simple Task		
Slave Strategy Master Input	OL Speed Position	CL Cartesian Position	CL Joint Position	OL Speed Position	CL Cartesian Position	CL Joint Position
Mean (w/ error)	81.637 151.029	82.321 110.255	88.138 184.342	79.01 92.577	68.196 88.591	78.57 153.863
Median (w/ error)	69.626 106.465	67.705 83.72	78.08 128.57	73.225 78.622	52.93 63.195	46.94 105.105
Min / Max (w/ error)	54.08/124.69 84.15/324.94	50.59/152.29 63.99/290.53	51.77/123.35 63.17/564.24	36.34/142.54 50.134/175.23	32.365/173.89 46.205/203.515	31.82/176.37 46.09/455.485
Best results (w/ error)	3 2	3 5	4 3	2 3	4 5	4 2

The backlash introduction gives the impression that nothing happens when one requests a direction change, so the user spontaneously moves further the interface. This situation could lead to the system instability because:

- in a speed control law, a large Cartesian speed reference is sent to the slave system, that reacts immediately when it comes out from the backlash zone;
- in a position control law, the error vector computed by the feedback loop increases rapidly during the backlash effect; the control loop tries to compensate this error, but the compensation can cause slave system oscillations.

For both situations, these behaviors are definitively unacceptable in a surgical scenario and, from a robotic point of view, they apply an important stress to the mechanical parts of the system. This points out the need to correctly take into account the major mechanical non-linearities. From these first tests one can see

that a Cartesian control with a position master input gives the best results for the majority of people involved in the simulations. The direct mapping between the instrument and the haptic interface workspaces, provided in this control strategy, gives a better sense of control for almost all users, because the position of the haptic interface directly represents the position of the instrument in the virtual scene, so the movement coordination is simpler. Also, the force feedback applied in the position input (described in section 3.3) helps to keep track between the interface and the instrument, especially when the instrument does not follow exactly the user requests because of a simulated error. On the real system, this means that an external tracking system is mandatory to correctly reproduce the force feedback effect, because the estimated instrument Cartesian position, computed with the theoretical kinematic model and based on the encoders positions, does not represent the actual instrument position. Finally, a direct mapping decreases the resolution of the movement because of the different sizes of instrument and interface workspaces. This problem could be resolved by scaling the interface movements with a filter on the user input: in this case it is necessary to implement a repositioning system (declutching) for the interfaces.

5   Conclusion

Simulation is an important tool in assessing the behavior of a robotic system, because it allows to create multiple testing sessions and to try different strategies before implementing them on the real system. But, to be useful, the simulator must closely replicate the system architecture, including the non-linearities that could influence the system control. The simulator presented in this paper, developed specifically for our novel flexible robot for no-scar surgery, has permitted us to conduct a preliminary test phase to assess several control strategies and the effect of mechanical non-linearities both on the system behavior and on the user response. Our first tests show that, with an ideal model, speed control permits to obtain better accuracy, while position control assures a faster system response. On the contrary, when the mechanical non-linearities are introduced inside the simulation, a position control strategy on the slave side, together with a direct mapping between the haptic interfaces and the instruments workspaces on the master side, permits to better compensate the non-linearities effects for the majority of people involved in the simulations. The force feedback applied on the haptic interfaces seems to play an important role in this case, because permits to keep track between the interface and the instrument positions, thus avoiding large errors in the control scheme. Transferring these conclusions on the real system, this points out the need to estimate the instrument Cartesian position with an external tracker, rather than rely on the encoders position, and to better compute a model that takes account of the non-linearities effects. We will next organize a more complete test session with both surgeons and non-surgeons users with a new simulator that will reproduce the physics interactions between virtual objects. The aim is to assess the control strategies for a specific surgical task that involves both arms, such as grab a tissue with one instrument and cut it with the other.

References

1. Whiteford, M.H., Swanstrom, L.L.: Emerging Technologies Including Robotics and Natural Orifice Transluminal Endoscopic Surgery (NOTES) Colorectal Surgery. Journal of Surgical Oncology 96, 678–683 (2007)
2. Romanelli, J.R., Earle, D.B.: Single-port laparoscopic surgery: an overview. Surg. Endoscop. 23, 1419–1427 (2009)
3. Webster III, R.J., Jones, B.A.: Design and Kinematic Modeling of Constant Curvature Continuum Robots: A Review. The International Journal of Robotics Research 29(13), 1661–1683 (2010)
4. Baerlocher, P., Boulic, R.: An inverse kinematics architecture enforcing an arbitrary number of strict priority levels. The Visual Computer 20, 402–417 (2004)
5. Allemann, P., Leroy, J., Asakuma, M., Al Abeidi, F., Dallemagne, B., Marescaux, J.: Robotics Overcome Technical Limitations of Single-Trocar Surgery. Arch. Surg. 145(3), 267–271 (2010)
6. Hannan, M.W., Walker, I.D.: Kinematics and the Implementation of an Elephant's Trunk Manipulator and other Continuum Style Robots. Journal of Robotic Systems 20(2), 45–63 (2003)
7. Jones, B.A., Walker, I.D.: Kinematics for Multisection Continuum Robots. IEEE Transactions on Robotics 22(1), 43–55 (2006)
8. Hanafusa, H., Yoshikawa, T., Nakamura, Y.: Analysis and control of articulated robot arms with redundancy. Preprints 8th Triennial IFAC World Congress, vol. 14, pp. 78–83 (1981)
9. Bardou, B., Nageotte, F., Zanne, P., de Mathelin, M.: Design of a telemanipulated system for transluminal surgery. In: IEEE 31st Int. Conf. on Engineering in Medicine and Biology, Minneapolis, Minnesota, September 2-6, pp. 5577–5582 (2009)
10. Bardou, B., Nageotte, F., Zanne, P., de Mathelin, M.: Improvements in the Control of a Flexible Endoscopic System. In: International Conference on Robotics and Automation, Saint Paul (MN) (May 2012)

Traceable Particle Swarm Optimization for Electromagnetically Navigated Bronchoscopy

Xiongbiao Luo[1], Takayuki Kitasaka[2], and Kensaku Mori[1,3]

[1] Information and Communications Headquarters, Nagoya University, Japan
[2] Faculty of Information Science, Aichi Institute of Technology, Japan
[3] Graduate School of Information Science, Nagoya University, Japan
xiongbiao.luo@gmail.com,
kitasaka@aitech.ac.jp,
kensaku@is.nagoya-u.ac.jp

Abstract. This paper proposes a modified evolutionary algorithm called traceable particle swarm optimization (PSO) that boosts bronchoscope motion tracking during electromagnetically navigated bronchoscopy. Since electromagnetic (EM) tracking is usually deteriorated by uncertainties (e.g., patient respiratory motion or magnetic field distortion) that occur in interventions, we develop a traceable PSO framework by integrating EM sensor measurements and image intensity information into the standard PSO method. In particular, all evolutionary parameters in our PSO framework can be updated traceably or adaptively in accordance with spatial distance constraints and image similarity information, resulting in an advantageous performance in dynamic bronchoscope motion estimation. Experimental results based on dynamic phantom validation demonstrate that our proposed tracking scheme provides a more robust, accurate, and efficient approach for endoscope motion tracking than several current available methods. The average tracking accuracy of position and orientation was improved from (4.3 mm, 7.8°) to (3.3 mm, 6.5°) and the computational time was reduced from 1.0 to 0.8 seconds per frame without any acceleration devices or code optimization strategy.

Keywords: Bronchoscope Motion Tracking, Electromagnetic Tracking, Particle Swarm Optimization, Electromagnetically Navigated Bronchoscopy.

1 Introduction

Endoscope location estimation or its motion tracking is the key component of any endoscope navigation systems, for instance, bronchoscope, colonscope, conchoscope, and neuroendoscope. Such a motion tracking procedure is usually formulated as an optimization process, which is commonly solved by deterministic [1–5, 9] or stochastic [6–8] approaches. Deterministic methods, typically intensity-based registration algorithms [3], usually define an optimization function to minimize the pixel difference between real video images and virtual bronchial renderings generated from pre-operative imaging information, for

example, three-dimensional (3-D) data that are acquired by computed tomography (CT) or magnetic resonance imaging (MRI) scanners. Although image-based methods work well in bronchoscope motion tracking, they are somewhat constrained by bronchoscopic image artifacts (e.g., motion blurring) and easily get trapped in local minima during optimization. On the other hand, since EM trackers suffer from localization problems (e.g., patient airway deformation) and inaccurate EM sensor measurements (e.g., magnetic field distortion due to metallic materials in the working volume), stochastic methods, which take the randomness of bronchoscope movements into account, were introduced to deal with the dynamic uncertainties in bronchoscope motion tracking. Such methods seek the optimal of the posterior probability of one bronchoscope motion state, e.g., using sequential Monte Carlo (SMC) algorithms to generate a set of particles and propagate them to approximate the probability distribution of dynamic states. From experimental results [6–8], stochastic approaches were proved to be stable and accurate. Compared to deterministic methods, stochastic approaches show more robust and precise tracking performance but require more computational time to estimate six degrees of freedom (6DoF) motion parameters.

Even though many papers have been published on stochastic methods for improving electromagnetically navigated bronchoscopy [6–8], a more robust and accurate optimization approach is still greatly expected to tackle stochastic ambiguities in bronchoscope motion tracking. Recently, a numerous population-based stochastic evolutionary algorithm, particle swarm optimization (PSO), which was originally proposed by Kennedy and Eberhart [10], has been increasingly applied as a successful optimization technique to address multidimensional complex problems [12, 11, 13]. The algorithm simulates natural and biological behaviors such as birds flocking and fish schooling to find optimal solutions in nonlinear and high-dimensional spaces. Moreover, one of most attractive aspects of PSO is that it can tackle nonlinear, non-differentiable, and multi-modal optimization problems by dynamically interacting all particles in a similar analogy with the "cognitive" and "social" properties of populations [15, 14].

This work develops a traceable PSO framework for boosting EM tracking during electromagnetically navigated bronchoscopy. It is worthwhile to highlight the following aspects of our proposed approach. First, to the best of our knowledge, our proposed PSO framework is a novel application of PSO in endoscope motion tracking. We successfully formulated endoscope motion tracking as a PSO-based stochastic optimization process. Video image information and EM sensor measurements can be effectively integrated into PSO to achieve a robust and accurate tracking method, which also provides an effective means to fuse other external tracking sources in bronchoscopy navigation. Furthermore, using spatial constraints and image similarity, we modified PSO to automatically refresh evolutionary parameters for addressing the diversity loss problem, alleviating particle impoverishment, and obtaining various particle diversity in PSO iterations. Finally, our proposed approach combined ideas from evolutionary computation and medical image computing communities that should be applicable to other endoscopic guidances, e.g., conchoscope or colonscope.

This paper is generally organized as follows. We briefly review the basic PSO algorithm in Section 2. In Section 3, our proposed method for bronchoscope location or navigation is described in detail, by following our validation setups that are presented in Section 4. Experimental results are shown in 5 and discussed in 7 before concluding this work and giving future work in Section 7.

2   Particle Swarm Optimization

We here briefly review the standard PSO algorithm [10]. In PSO, a number of particles are utilized to denote the solutions in a dynamic system. Each particle i at iteration j and time k is represented by state vector $\mathbf{x}_k^{i,j} \in \Re^D$ associated with velocity vector $\mathbf{v}_k^{i,j} \in \Re^D$ that conducts particle transition and a corresponding fitness value that is determined by observation model $f(\mathbf{x}_k^{i,j})$. Given a particle set $\{\mathbf{x}_k^{i,j} \in \Re^D\}_{i=1}^N$ (N is the number of particles), in j-th iteration, particle state $\mathbf{x}_k^{i,j}$ and its velocity vector $\mathbf{v}_k^{i,j}$ are propagated to $\mathbf{x}_k^{i,j+1}$ and $\mathbf{v}_k^{i,j+1}$ with inertia weight ω (to decide how much $\mathbf{v}_k^{i,j}$ to be preserved in $\mathbf{v}_k^{i,j+1}$) by:

$$\mathbf{v}_k^{i,j+1} = \omega\mathbf{v}_k^{i,j} + \lambda_1\eta_1(\mathbf{p}_k^{i,j} - \mathbf{x}_k^{i,j}) + \lambda_2\eta_2(\mathbf{g}_k^{i,j} - \mathbf{x}_k^{i,j}), \tag{1}$$

$$\mathbf{x}_k^{i,j+1} = \mathbf{x}_k^{i,j} + \mathbf{v}_k^{i,j+1}, \tag{2}$$

where λ_1 and λ_2 are acceleration constants and η_1 and η_2 are randomly generated from the uniform distribution with interval [0.0 1.0]. $\mathbf{p}_k^{i,j}$ (for the local individual best) and $\mathbf{g}_k^{i,j}$ (for the global all best) are the best state found by particle i so far and the best state found by the whole swarm so far, respectively.

After j-th iteration, $\mathbf{p}_k^{i,j}$ and $\mathbf{g}_k^{i,j}$ can be updated in accordance with each particle fitness value evaluated by $f(\mathbf{x}_k^{i,j+1})$:

$$\mathbf{p}_k^{i,j+1} = \begin{cases} \mathbf{x}_k^{i,j+1} & \text{if} f(\mathbf{x}_k^{i,j+1}) > f(\mathbf{p}_k^{i,j}) \\ \mathbf{p}_k^{i,j} & \text{otherwise} \end{cases}, \tag{3}$$

$$\mathbf{g}_k^{i,j+1} = \arg max_{\mathbf{p}_k^{i,j+1}} f(\mathbf{p}_k^{i,j+1}). \tag{4}$$

Based on Eqs. 1∼4, PSO tries to find the optimal solution during an optimization procedure. Please refer to [10] for more details about the basic PSO algorithm.

3   Proposed Tracking Framework

3.1   Overview

Our proposed framework to estimate bronchoscope motion consists of three main steps: (1) particle stochastic diffusion, (2) traceable analysis of evolutionary factors during particle propagation, and (3) the determination of bronchoscope motion parameters. During Step (1), a swarm of particles is generated and initialized. These particles are randomly propagated to increase the diversity.
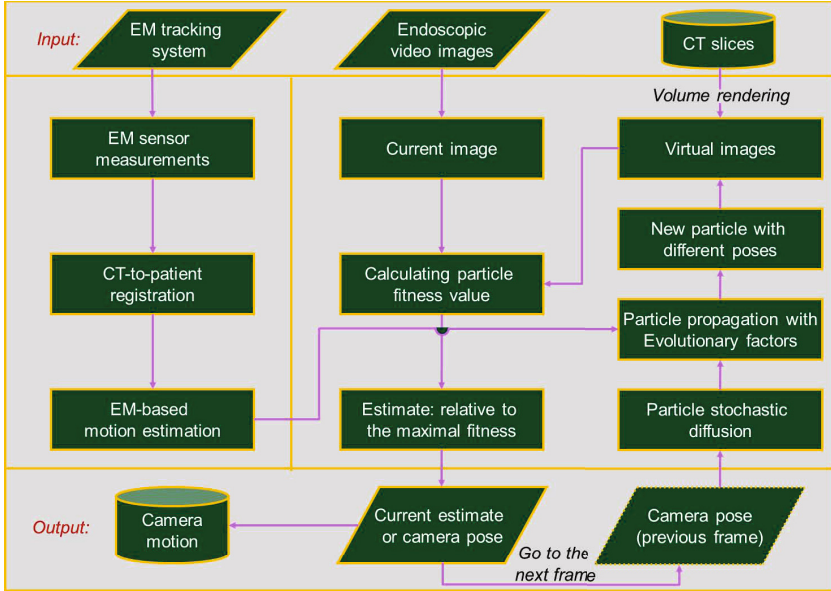
Fig. 1. The flowchart of bronchoscope 3-D motion estimation using our proposed method that comprises three steps of stochastic diffusion, particle propagation with adaptive evolutionary factors, and the determination of pose parameters

After that, evolutionary parameters including ω, λ_1, and λ_2 in Eqs. 1 and 2 are calculated in Step (2). Finally, PSO iterations are performed to determine the bronchoscope motion parameters or its pose information with position and orientation. Fig. 1 illustrates the proposed tracking framework to process CT slices, endoscopic video images, and EM sensor measurements for motion estimation.

3.2   Particle Stochastic Diffusion

We first define particle state $\mathbf{p}_k^{i,j}$ as a six-dimensional vector based on the bronchoscope position and orientation in our case of bronchoscope motion tracking:

$$\mathbf{p}_k^{i,j} = [t_x \quad t_y \quad t_z \quad \theta \quad \phi \quad \psi]^T, \tag{5}$$

where $\mathbf{p}_k^{i,j}$ corresponds to camera pose matrix $\mathbf{Q}_k(\mathbf{p}_k^{i,j}) = \mathcal{F}(t_x, t_y, t_z; \theta, \phi, \psi)$, translations t_x, t_y, and t_z and Euler angles θ, ϕ, and ψ of the bronchoscope camera around the x-, y-, and z-axes represent position and rotation, respectively.

Suppose that we generate a swarm of particles $\mathcal{P}_k^{i,j} = \{(\mathbf{p}_k^{i,j}, f(\mathbf{p}_k^{i,j}), \gamma_k^{i,j})\}_{i=1}^N$, where $\gamma_k^{i,j}$ is a weight based on spatial distance constraints. To increase the diversity of particles and avoid particle impoverishment, we perform a stochastic diffusion procedure in terms of the Gaussian propagation model and obtain $\mathbf{x}_k^{i,j}$:

$$\mathbf{x}_k^{i,j} = \mathcal{G}(\mathbf{p}_k^{i,j}, \mu \Delta \mathbf{s}_k), \tag{6}$$

where μ is a Gaussian distribution random number: $r \sim \mathcal{N}(0,1)$ and $\Delta \mathbf{s}_k$ is determined by EM-based motion estimates $\mathbf{s}_k$ and $\mathbf{s}_{k-1}$ at frames k and $(k-1)$:

$$\Delta \mathbf{s}_k = \mathbf{s}_k - \mathbf{s}_{k-1}, \mathbf{s}_k = [t_x^k \; t_y^k \; t_z^k \; \theta^k \; \phi^k \; \psi^k]_{EM}^T, \qquad (7)$$

which is also used for initializing transition velocity $\mathbf{v}_k$: $\mathbf{v}_k = \Delta \mathbf{s}_k$ before it is updated by the global best solutions or estimates $\mathbf{g}_k$ and $\mathbf{g}_{k-1}$ during the iterations:

$$\mathbf{v}_k = \mathbf{g}_k - \mathbf{g}_{k-1}, \mathbf{g}_k = [t_x^k \; t_y^k \; t_z^k \; \theta^k \; \phi^k \; \psi^k]_{global}^T. \qquad (8)$$

Note that our stochastic diffusion procedure for particle diversification does not perform a *resampling process*, as SMC or particle filter methods do [8], since the local best particles provide compact samples for propagation [13].

3.3   Parameter Traceable Analysis

Evolutionary parameters λ_1, λ_2, and ω heavily influence the PSO performance. Most current modified PSO algorithms do not consider spatial continuity constraint and image sequence information, which may result in a lack of systematic treatment of evolutionary states and expose PSO to a dangerous level of swarm explosion and divergence. To handle this limitation, we modify PSO based on image intensity to traceably control λ_1, and λ_2 by:

$$\lambda_1 = 2f(\mathbf{p}_k^{i,j})/f(\mathbf{p}_k^{i,j}) + f(\mathbf{g}_k^{i,j}), \quad \lambda_2 = 2f(\mathbf{g}_k^{i,j})/f(\mathbf{p}_k^{i,j}) + f(\mathbf{g}_k^{i,j}), \qquad (9)$$

where $f(\mathbf{p}_k^{i,j})$ is defined as observation probability $Pr(\mathbf{o}_k^{i,j}|\mathbf{p}_k^{i,j})$:

$$f(\mathbf{p}_k^{i,j}) = Pr(\mathbf{o}_k^{i,j}|\mathbf{p}_k^{i,j}) = \delta_k^{i,j}(\sum_{i=1}^{N} \delta_k^{i,j})^{-1}, \qquad (10)$$

where $\mathbf{o}_k^{i,j}$ is an observation corresponding to $\mathbf{p}_k^{i,j}$. $Pr(\mathbf{o}_k^{i,j}|\mathbf{p}_k^{i,j})$ depends on similarity $\delta_k^{i,j}$ between video image I_R^k and virtual rendering $I_V^k(\mathbf{Q}_k)$ generated at pose matrix $\mathbf{Q}_k(\mathbf{p}_k^{i,j})$ and $\delta_k^{i,j}$ is calculated based on image intensity by a modified mean square error ($MoMSE$) [8]:

$$\delta_k^{i,j} = MoMSE(I_R^k, I_V^k(\mathbf{p}_k^{i,j})). \qquad (11)$$

For adaptively calculating ω, we utilize both fitness value $f(\mathbf{x}_k^{i,j}) \in [0 \; 1]$ and particle spatial distribution information $\gamma_k^{i,j}$ among the particles. We first compute average distance $d_k^{i,j}$ from one particle to all other particles:

$$d_k^{i,j} = \frac{1}{N-1} \sum_{i=1,i\neq t}^{N} \sqrt{(\mathbf{x}_k^{i,j} - \mathbf{x}_k^{t,j})^2}. \qquad (12)$$

After finding the largest distance d_{max} and the smallest distance d_{min} from $\{d_k^{i,j}\}_{i=1}^{N}$, we normalize distance $d_g^{i,j}$ between one particle and the current global best particle and obtain $\gamma_k^{i,j}$ and assign it to each particle:

$$\gamma_k^{i,j} = (d_g^{i,j} - d_{min})/(d_{max} - d_{min}), \quad \gamma_k^{i,j} \in [0 \; 1]. \qquad (13)$$

Finally, since ω was suggested within the interval [0.4 0.9] for weighting the global and the local searching abilities [14], we can traceably calculate it by:

$$\omega(f(\mathbf{x}_k^{i,j}), \gamma_k^{i,j}) = \frac{2}{2 + 3\exp(-1.28(f(\mathbf{x}_k^{i,j}) + \gamma_k^{i,j}))}, \tag{14}$$

which shows a novel strategy to automatically control ω in our modified PSO.

3.4 Bronchoscope Motion Estimation

Our work is to estimate to a full 6 degrees of freedom camera motion matrix $\mathbf{Q}_k$ including camera position and orientation. We integrate EM sensor measurements and image similarity information into our modified PSO algorithm discussed above. The output of the POS tracking framework is represented by:

$$\mathbf{Q}_k^*(g_k^*) = \mathcal{F}(t_x^g, t_y^g, t_z^g; \theta^g, \phi^g, \psi^g) = \begin{pmatrix} \mathbf{R}_k & \mathbf{t}_k \\ \mathbf{0}^T & 1 \end{pmatrix}, \tag{15}$$

where $\mathbf{t}_k = [t_x^g, t_y^g, t_z^g]^T$ and rotation matrix $\mathbf{R}_k$ are related to θ^g, ϕ^g, and ψ^g:

$$\mathbf{R}_k = \begin{pmatrix} C_3 C_2 & C_3 S_2 S_1 - S_3 C_1 & C_3 S_2 C_1 + S_3 S_1 \\ S_3 C_2 & S_3 S_2 S_1 + C_3 C_1 & S_3 S_2 C_1 - C_3 S_1 \\ -S_2 & C_2 S_1 & C_2 C_1 \end{pmatrix}, \tag{16}$$

where $S_1 = \sin\theta^g$, $S_2 = \sin\phi^g$, $S_3 = \sin\psi^g$, $C_1 = \cos\theta^g$, $C_2 = \cos\phi^g$, and $C_3 = \cos\psi^g$. The implementation of our proposed method for improving EM tracking and boosting navigated bronchoscopy is summarized in Algorithm 1.

Algorithm 1. Traceable PSO for Bronchoscope Motion Estimation

---

    input : Bronchoscopic video images $\mathbf{I}_R^k$, CT-based virtual images $\mathbf{I}_V$, and electromagnetic sensor measurements $\mathbf{s}_k$

    output: All global best estimates $\mathbf{Q}_k^*(g_k^*)$ of bronchoscope camera poses

    1. Initialization: At time $k = 0$, use $\mathbf{Q}_0$ to initialize $\mathcal{P}_0^{i,j}$ and $\mathbf{g}_0^{i,j}$;

    2. Perform stochastic diffusion to obtain $\{\mathbf{x}_0^{i,j}\}$ by Eq. 6;

    3. Compute $f(\mathbf{p}_0^{i,j})$ by Eq. 10 and update $\mathcal{P}_0^{i,j}$ and $\mathbf{g}_0^{i,j}$ by Eqs. 3 and 4;

    4. Implement Traceable PSO iterations:

    for $k = 1$ **to** T (frame number) **do**

        for $j = 1$ **to** M (iteration number) **do**

            Update evolutionary parameters ω, λ_1, and λ_2 by Eqs. 9 and 14;

            for $i = 1$ **to** N (particle number) **do**

                Perform PSO iteration in accordance with Eqs. 1 and 2;

                Update each particle $\mathbf{x}_k^{i,j}$ and velocity vector $\mathbf{v}_k^{i,j}$;

                Compute $f(\mathbf{p}_k^{i,j})$ and $\gamma_k^{i,j}$ by Eqs. 10 and 13;

            end

            Update particle set $\mathcal{P}_k^{i,j}$ and global best particle $\mathbf{g}_k^{i,j}$ by Eqs. 3 and 4;

        end

        Find global best estimate g_k^* from particle set $\mathcal{P}_k^{i,j}$;

        Determine motion pose: $\mathbf{Q}_k^*(g_k^*) \iff \mathcal{F}(t_x^g, t_y^g, t_z^g; \theta^g, \phi^g, \psi^g)$ by Eq. 15;

    end

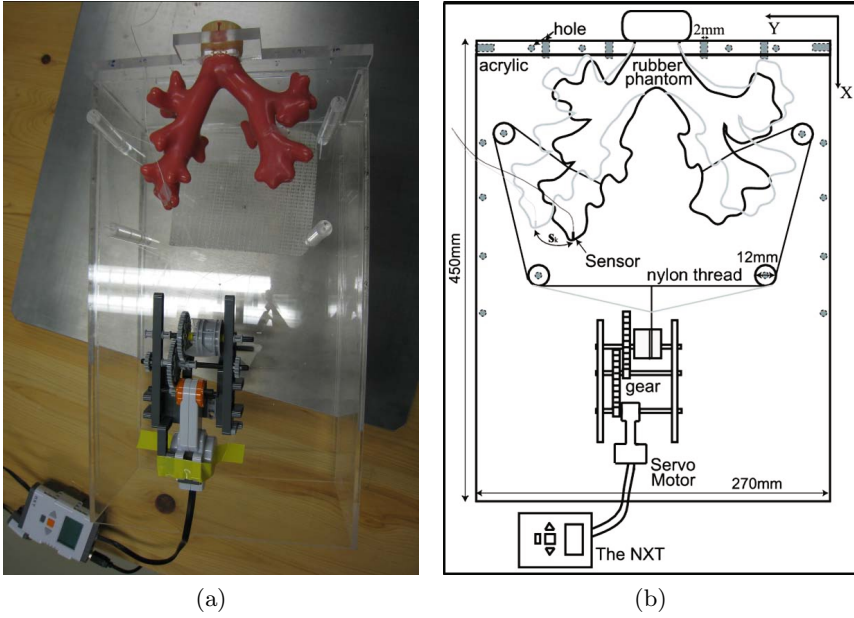   5. Return: all $\{\mathbf{Q}_k^*(g_k^*)\}_{k=1}^T$

Fig. 2. The dynamic phantom was constructed with the airway tree rubber, a motor, and nylon thread. (a) physical phantom and (b) phantom movement.

Fig. 3. The EM tracking system that was used here includes the control unit (*left*) and the fat-type magnetic field generator (*right*)

4   Experimental Setups

4.1   Hardwares

We evaluated our proposed tracking method on a dynamic phantom with an adjustable motion: $0 \sim 24$ mm, as shown in Fig 2. The CT spacing parameters of our phantom were: 512×512 pixels, 1021 slices, 0.68-mm reconstruction pitch, and 0.5-mm thick slices. A 3-D Guidance medSAFE tracker (Ascension Technology Corporation, USA) was used as an EM tracking system, which includes a 9-coil at flat-type transmitter as a magnetic field generator, as illustrated in Fig. 3. Endoscopic video images of size 362×370 pixels were recorded at 30 frames per second using an endoscope (BF-P260F, Olympus, Tokyo).

4.2   Ground Truth

To evaluate the tracking accuracy of different methods, we generated five ground truth datasets (GTDs) by manually adjusting the position and orientation of the virtual camera to qualitatively register the real and virtual bronchoscopic viewing points by hand. Three observers of the authors independently and repeatedly collected these GTDs in multiple sessions. We clarify that intra-observer consistency was 1.81 mm and 5.9°, 1.76 mm and 4.9°, and 1.93 mm and 4.8° from three observers, respectively; inter-observer consistency was 1.71 mm and 5.6°. Note that the clinical requirement of position and orientation is below 2 mm and 6° during bronchoscopic interventions.

   We compared five tracking approaches (1) M1: only using EMT tracking reported by Schwarz et al. [16], (2) M2: a hybrid method presented by Mori et al. [4], (3) M3: a modified hybrid method proposed by Luo et al. [17], (4) M4: a SMC-based solution introudced by Luo et al. [8], and (5) our proposed framework, as discussed in Section 3. Additionally, we set the particle number: $N = 50$ and the iteration number: $M = 10$. During the SMC-based tracking, the particle number was set to 500. We have done all implementations on a Microsoft Visual C++ platform and ran it on a conventional PC (CPU: Intel(R) Xeon(R) X5482 ×2 processors, 16-GByte main memory).

5   Results

Table 1 displays the quantitative results of the tracking accuracy from different methods. The average position and orientation errors of the proposed framework were 3.3 mm and 6.5°, which are definitely better than those of the previous published methods that had average errors of at least 4.3 mm and 7.8°. We also visually inspected the tracking results by manually checking whether the real images resembled the virtual images. Fig. 4 shows examples of real images and the corresponding virtual images generated from the camera parameters

Table 1. Quantitative results of tracking accuracy of compared methods in terms of position and orientation errors between estimated results and ground truth

| Experiments	Data 1	Data 2	Data 3	Data 4	Data 5
Max Motion	2.5 mm	5.6 mm	10.4 mm	13.8 mm	22.3 mm
M1 [16]	4.2±2.6 mm 6.7±5.2°	5.3±3.5 mm 8.8±6.2°	5.6±2.8 mm 7.9±5.3°	6.0±2.6 mm 9.6±6.0°	7.2±3.5 mm 13.5±11.1°
M2 [4]	3.8±3.2 mm 6.1±4.1°	4.9±4.2 mm 7.6±5.5°	5.4±3.2 mm 6.8±6.2°	5.8±3.6 mm 8.8±5.6°	6.8±4.4 mm 12.9±13.4°
M3 [17]	3.4±2.6 mm 5.3±3.2°	4.6±3.5 mm 6.7±2.9°	5.3±4.1 mm 5.6±5.2°	5.6±4.8 mm 10.6±5.8°	6.1±4.6 mm 12.7±11.8°
M4 [8]	3.1±2.2 mm 4.8±4.2°	3.9±2.1 mm 5.8±3.2°	4.1±2.5 mm 6.2±3.1°	4.6±3.2 mm 9.5±5.5°	5.6±4.3 mm 12.9±12.6°
Our method	2.6±2.4 mm 3.9±2.2°	2.9±1.9 mm 4.2±2.6°	3.2±2.6 mm 5.2±3.6°	3.5±2.9 mm 8.9±5.2°	4.4±3.0 mm 10.2±10.5°
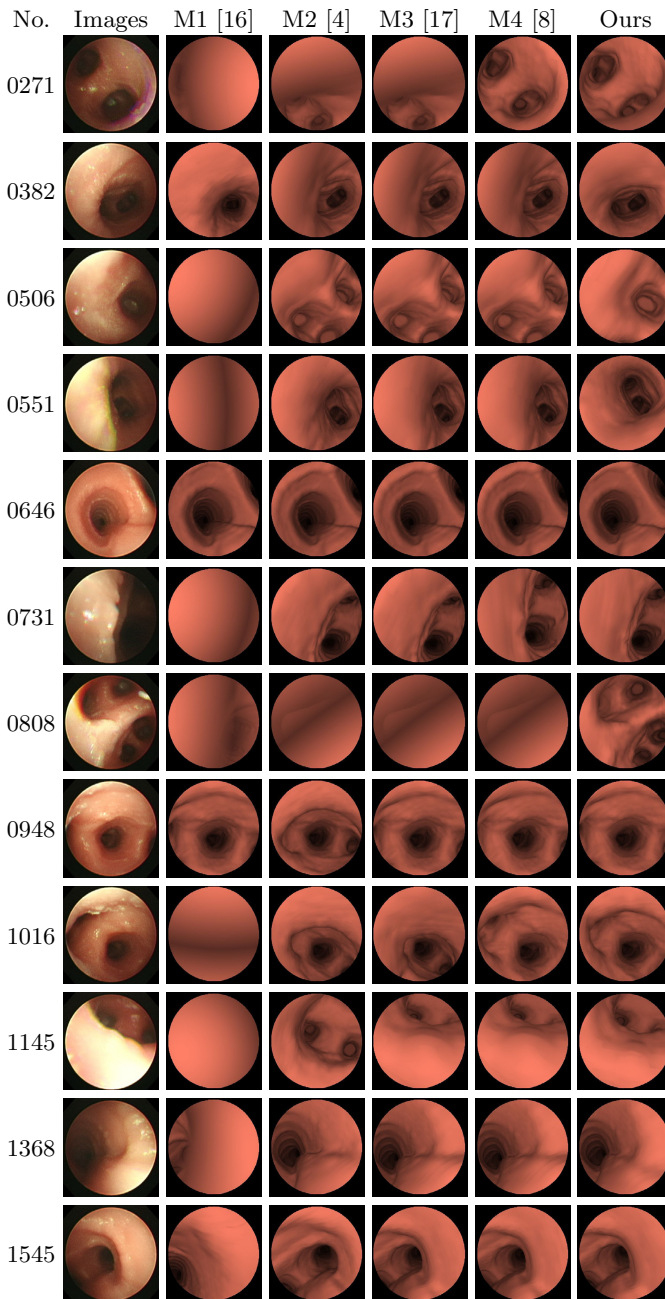
Fig. 4. Visual comparison of tracking results of Data 4. Left column shows selected frame numbers, and second column gives their corresponding video images. Other columns display virtual bronchoscopic images generated from tracking results using methods discussed above. Our method shows the best performance.

estimated by each method. This visual investigation of the successfully processed frames further demonstrates the effectiveness of our proposed method. Additionally, the current runtime of our method is about 0.8 seconds per frame, which outperforms that reported in [5] by speed-up devices (0.98 seconds).

6   Discussion

Generally, our proposed PSO tracking method provides a more accurate and robust strategy to estimate endoscope motion than previous approaches. We attribute such an advantageous performance of our PSO framework tracking to the following aspects. First, we believe that our traceable PSO is partly an association of PSO iterations and SMC sampling procedures, and hence it outperforms the SMC sampling algorithms in motion tracking. During SMC sampling procedures, a successful particle sampling depends heavily on the proposal distribution function [18]. Particles with large weights located in the useful area of the proposal distribution are usually sampled. In fact, the proposal distribution is suggested to be the dynamic transition distribution, which may incur particles with larger weight that are not sampled when the useful area of the transition distribution stays at the tail of the observation distribution [18]. However, the PSO framework performs more like a hierarchical sampling strategy which propagates the particles integrated with the newest observations [11], possibly resolving the particle impoverishment problem. Next, automatically or traceably controlling evolutionary parameters is greatly helpful to update particles in iterations. The two acceleration factors, which were calculated based on the fitness value from the image intensity information, are more reasonable than setting them to 2 in standard PSOs [10]. Moreover, the inertia weight is also adaptively determined by both spatial distance constraint and image intensity information, resulting in more flexibly balancing the global and local search abilities and providing a reasonable velocity limitation to move particles. Finally, without any *resampling process* in our method, compared to SMC sampling or particle filtering, it is helpful to reduce the runtime of our method.

Additionally, we must clarify the potential limitations of our proposed methods. Particle robustness, which means the particle fitness value to be correctly computed and evaluate, depends somewhat on image intensity information. However, the image artifacts that occur in bronchoscopic video may collapse the correct computation of the fitness value. To tackle this drawback, a more robust intensity similarity measure, which should be slightly insensitive to illumination changes or other image artifacts, must be developed in future work. Another problem remains how to properly choose the particle and iteration numbers M and N. In fact, it is difficult to know their influences on the tracking performance. Thoroughly evaluating M and N in PSO is another future work. We plan to adaptively select them by particle robustness during iterations, possibly reducing our current runtime, which is also future work.

7   Conclusions

This work proposes a new bronchoscope 3-D motion tracking framework for endoscope location or navigation using a traceable PSO algorithm that can refresh its evolutionary parameters based on spatial distance constraints and image intensity information during iterations. Dynamic phantom validation proves that our method provides a more advantageous tracking performance than state-of-the-art methods. Future work also includes reducing the runtime of our method.

References

1. Deligianni, F., Chung, A.J., Yang, G.Z.: Nonrigid 2-D/3-D registration for patient specific bronchoscopy simulation with statistical shape modeling: Phantom validation. IEEE TMI 25(11), 1462–1471 (2006)
2. Helferty, J.P., Sherbondy, A.J., Kiraly, A.P., Higgins, W.E.: Computer-based system for the virtual-endoscopic guidance of bronchoscopy. CVIU 108(1-2), 171–187 (2007)
3. Deguchi, D., Mori, K., Feuerstein, M., Kitasaka, T., Maurer Jr., C.R., Suenaga, Y., Takabatake, H., Mori, M., Natori, H.: Selective image similarity measure for bronchoscope tracking based on image registration. MedIA 13(4), 621–633 (2009)
4. Mori, K., Deguchi, D., Akiyama, K., Kitasaka, T., Maurer Jr., C.R., Suenaga, Y., Takabatake, H., Mori, M., Natori, H.: Hybrid bronchoscope tracking using a magnetic tracking sensor and image registration. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005, Part II. LNCS, vol. 3750, pp. 543–550. Springer, Heidelberg (2005)
5. Reichl, T., Luo, X., Menzel, M., Hautmann, H., Mori, K., Navab, N.: Deformable registration of bronchoscopic video sequences to CT volumes with guaranteed smooth output. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part I. LNCS, vol. 6891, pp. 17–24. Springer, Heidelberg (2011)
6. Gergel, I., dos Santos, T.R., Tetzlaff, R., Maier-Hein, L., Meinzer, H.-P., Wegner, I.: Particle filtering for respiratory motion compensation during navigated bronchoscopy. In: Wong, K.H., Miga, M.I. (eds.) SPIE Medical Imaging 2010, California, USA, vol. 7625, p. 76250W (2010)
7. Soper, T.D., Haynor, D.R., Glenny, R.W., Seibel, E.J.: In vivo validation of a hybrid tracking system for navigation of an ultrathin bronchoscope within peripheral airways. IEEE TBME 57(3), 736–745 (2010)
8. Luó, X., Reichl, T., Feuerstein, M., Kitasaka, T., Mori, K.: Modified hybrid bronchoscope tracking based on sequential monte carlo sampler: Dynamic phantom validation. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) ACCV 2010, Part III. LNCS, vol. 6494, pp. 409–421. Springer, Heidelberg (2011)

9. Luo, X., Kitasaka, T., Mori, K.: Bronchoscopy navigation beyond electromagnetic tracking systems: A novel bronchoscope tracking prototype. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part I. LNCS, vol. 6891, pp. 194–202. Springer, Heidelberg (2011)
10. Kennedy, J., Eberhart, R.C.: Particle swarm optimization. In: Proc. IEEE International Conf. on Neural Networks, pp. 1942–1948 (1995)
11. Zhang, X., Hu, W., Qu, W., Maybank, S.: Multiple object tracking via species-based particle swarm optimization. IEEE TCSVT 20(11), 1590–1602 (2010)
12. John, V., Trucco, E., Ivekovic, S.: Markerless human articulated tracking using hierarchical particle swarm optimisation. IVC 28(11), 1530–1547 (2010)
13. Zhang, L., Mei, T., Liu, Y., Tao, D., Zhou, H.-Q.: Visual search reranking via adaptive particle swarm optimization. Pattern Recognition 44(8), 1811–1820 (2011)
14. Parrott, D., Li, X.: Locating and tracking multiple dynamic optima by a particle swarm model using speciation. IEEE TEC 10(4), 440–458 (2006)
15. Clerc, M., Kennedy, J.: The particle swarm-explosion, stability and convergence in a multidimensional complex space. IEEE TEC 6(1), 58–73 (2002)
16. Schwarz, Y., Greif, J., Becker, H.D., Ernst, A., Mehta, A.: Real-time electromagnetic navigation bronchoscopy to peripheral lung lesions using overlaid CT images: the first human study. Chest 129(4), 988–994 (2006)
17. Luo, X., Feuerstein, M., Imaizumi, K., Hasegawa, Y., Mori, K.: Towards hybrid bronchoscope tracking under respiratory motion: evaluation on a dynamic motion phantom. In: Wong, K.H., Miga, M.I. (eds.) SPIE Medical Imaging 2010, California USA, vol. 7625, p. 76250B (2010)
18. Arulampalam, M., Maskell, S., Gordon, N., Clapp, T.: A tutorial on particle filters for nonlinear/non-gaussian Bayesian tracking. IEEE TSP 50(2), 174–188 (2002)

Augmented 3-D View for Laparoscopy Surgery

Brahim Tamadazte, Sandrine Voros, Christophe Boschet, Philippe Cinquin,
and Céline Fouard

TIMC-IMAG Laboratory, UMR 5525 UJF CNRS, GMCAO Team,
Faculty of Medicine, 38706 La Tronche Cedex, France
{Brahim.Tamadazte,Sandrine.Voros,Christophe.Boschet,Philippe.Cinquin,
Celine.Fouard}@imag.fr
http://www-timc.imag.fr/

Abstract. Minimally invasive surgery became popular as it benefits the
patient, reducing pain and allowing better recovery. However, during
endoscopy, surgeons face significant technical challenges such as indirect
2-D visualization, restricted field-of-view, lack of brightness, insufficient
depth-of-field, distortions, etc. In this paper, we propose a new vision
system for laparoscopy. This system consists of two pairs of cameras in
stereoscopic conditions directly inserted into the patient through a single
standard 10 millimeters diameter trocar and be deployed and fixed into
the patient's abdominal cavity. It answers several problems raised by
standard endoscopic systems as well as a 3-D virtual environment.

Keywords: multiple views vision system, CMOS cameras, laparoscopy
surgery, 3-D visualization.

1 Introduction

In laparoscopic surgery, also called minimally invasive surgery (MIS), surgeons
access the abdominal cavity (or the rib cage) through small skin incisions. These
incisions are usually small with a diameter inferior to ten millimeters and allow
for the insertion of a rod lens system which is connected to a video camera,
a light source and a video monitor [15]. This approach offers decreased blood
loss and postoperative pain, in addition to less hemorrhaging, shorter recovery
time, and better cicatrization [9] than open surgery. However, the surgeon faces
physical and visual drawbacks such as described in Figure 1.

To overcome these drawbacks, other types of endoscopes/vision devices have
been explored in the literature (some are now commercialized) such as stereoen-
doscopes [4], or a digital camera directly placed inside the abdominal cavity [6].
Other solutions, based on software approaches have been proposed as for instance
augmented reality, where a virtual model of the operative site is projected on
the real image [1], [3], [7], or virtual reality where a 3-D virtual environment of
the surgical scene is displayed so that the surgeon can plan the surgery or be
guided in real time [11], [10]. However, registration between preoperative and
intraoperative images, especially in the case of soft tissue remains challenging.
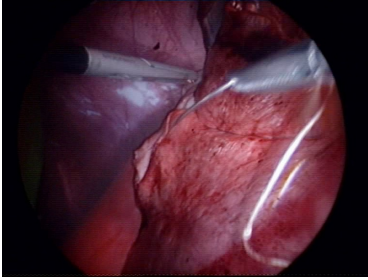
- no direct physical access to the internal organs (neither for visualization, nor for manipulation)
- 2-D information retrieved by a video camera
- a restricted field of view
- no access to hidden parts of organs
- specular reflections of the light
- image distorsions [14]

Fig. 1. Example of standard endoscopic image and associated limitations

In this paper, we propose a new endoscopic vision system. It consists of one or two stereoscopic device(s). Each device is made of two high resolution CMOS[1] video cameras rigidly fixed in stereoscopic conditions and covered by a polarized film. They are designed to be compatible with operating conditions (no toxic material, can be sterilized) and small enough to be introduced within the patient through a single 10 mm diameter trocar. Once in the patient's abdomen, they can be positioned and fixed by the surgeon thanks to a lasso system. Such a system will be able to solve several problems raised by standard endoscopic vision systems. Indeed even before looking into the benefits of stereoscopic systems, the simple 2-D images provided by each camera will offer a better resolution and illumination, less specular reflections, and the system in itself allows a better FOV (as there are several cameras carefully placed), and more degrees-of-freedom for the visualization. In addition the calibrated stereoscopic conditions allow for a 3-D reconstruction of the abdominal cavity and thus virtual movements of the camera without physically moving the system, as well as a possibility to create a virtual point of view that is not restricted by the fixed insertion point of the traditional endoscope.

This paper is structured as follows: section 2 presents the different materials used to develop the proposed multiple views vision device. Section 3 gives the methods used to perform camera calibration and compute the 3-D virtual environment. Section 4 illustrates the experimental setup used to validate the proposed materials and describes the obtained results such as the 3-D virtual environment computation and Section 5 presents a discussion about the obtained results and the improvements to be made in future developments.

2   Materials

This section presents the material used to develop a stereovision system for enhanced endoscopy that fully fulfills the conditions required for surgical procedures. Each material has thus been chosen not only for its high technological performance, but also to fit actual surgical constraints.

---

[1] Complementary Metal-Oxide-Semiconductor.

2.1   CMOS Cameras

The development of multiple views vision system includes four high performance and ultra small CMOS cameras developed by our partner ST Microelectronics® (see bottom right close-up view in Figure 4). Their sensors resolution is 2 megapixel (1600 × 1200 pixels) with a pixel's size of 1.75 μm × 1.75 μm. Their compact size (5 × 5 × 3.8 mm^3), their management of low illumination (≤ 10 lux), their high frame rate (30 frames/second) and their high quality lens (FOV of 51° of each camera) with low TV distortion (≤ 1%) make them perfectly suitable for laparoscopic conditions. The cost of the used cameras is about few US dollars (in large scale diffusion).

2.2   Polarized Films

When the light emitted from the optic fiber of a standard endoscope reflects on the organs and directly goes back to the camera with an angle of 180° it creates specular reflections. These reflections saturate the video cameras and create white areas in the images where visual information is lost. This phenomenon is amplified when the organs are covered with a liquid mix of several fluids.
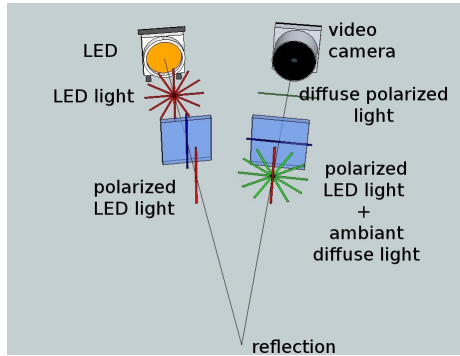

Fig. 2. Polarized system to resolve the problem of specular reflections using a polarized film on the camera oriented perpendicularly to a polarized film placed on the light source

There exist several ways to suppress specular reflections. Some are algorithmic [12] and [8]. Here, we choose a mechanical solution by placing polarized films in front of the light and in front of cameras as illustrated Figure 2. The obtained results are presented in section 4.

2.3   Stereoscopic Device

Two hermetic stereoscopic *capsules* were designed and manufactured. For each capsule, two video cameras (Figure 3 (a)) are positioned in stereoscopic conditions with an angle of 6°. They are rigidly fixed to each other so that they can be

calibrated (Figure 3 (b)). A linear polarized film is placed in front of each video camera, and the whole system is shielded with a biocompatible resin in order to insure water resistance (Figure 3 (c)). The obtained capsule has a diameter inferior to 10 mm and can be inserted with another one into a single trocar.
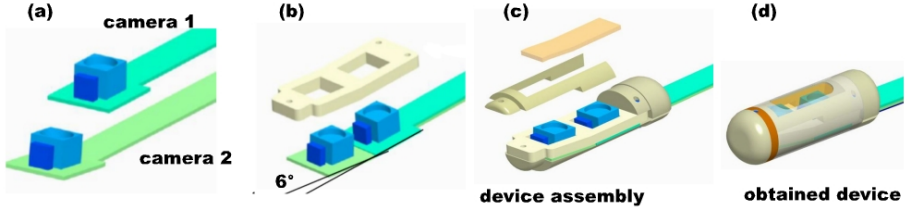

Fig. 3. Computer aided design (CAD) model showing different steps of the *capsules* composition

2.4 Deployment System Design

A deployment and extraction system is proposed to insert, fix and extract the stereoscopic devices. This concept is based on the use of a lasso system allowing to catch the capsules tip and able to fix them (i.e. stabilize the cameras position in the abdominal cavity). Two additional 5 mm diameter incisions are necessary to introduce the lassos within the abdomen in order to catch the both capsules as illustrated in figure 4. Each capsule is also equipped with a small semi-rigid line allowing to remove the vision system through the trocar. The extraction is guided by the conical and rounded shape of the capsule ends (see Figure 3.(d)). The final version of the developed multiple views system is shown in Figure 4.

The configuration of the proposed multiple views vision system when it is deployed inside the patient's abdominal cavity is shown in Figure 5.

3 Methods

During a surgical intervention, the surgeon uses an indirect vision, and thus loses 3D information. Our system proposes a 3D reconstruction of the operating site, as realistic as possible to allow the surgeon to access distance, size and texture information.

3.1 Cameras Calibration

To take the best advantages of a stereoscopic device, it is necessary to be able to model the vision systems. The chosen model for video cameras is projective. To do, we use Zhang's method [16] calibrate the cameras and estimate: 1) the
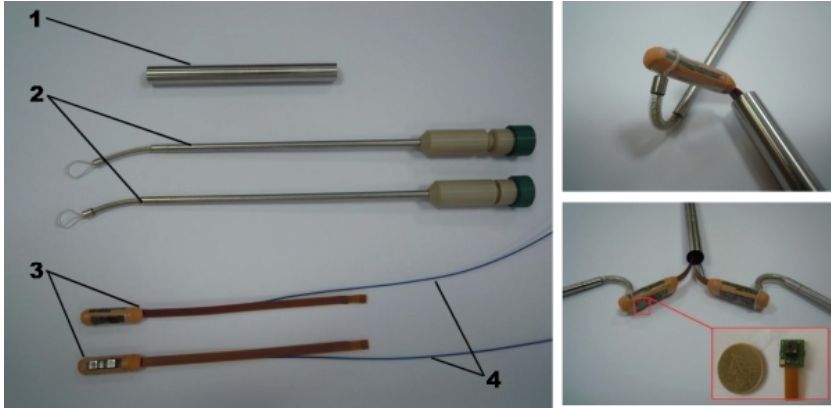
Fig. 4. Pictures which represent the developed multiple views vision system concepts: (1) is a 10 mm diameter trocar, (2) are lassos allowing stabilization of the stereoscopic devices inside the abdominal cavity, (3) are the proposed stereoscopic devices, and (4) are small semi-rigid line allowing the removal the vision systems
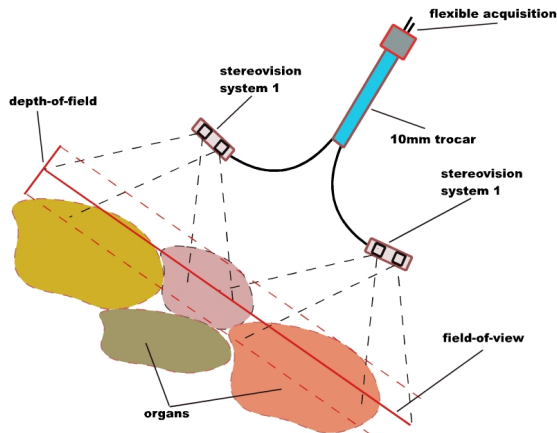

Fig. 5. Mapping of the cameras deployment inside the abdomen cavity and the benefits of the developed system in terms of FOV and DOF

intrinsic parameters of each camera, 2) the distortions and 3) the rigid transformation matrix between the different cameras of a pair. To compute the intrinsic parameters of each camera, a 2-D checkerboard pattern is viewed from several orientations. After this first calibration step, it is necessary to know the relative position of one camera with respect to any other by computing the interest points (using Harris detector [5]) and their correspondence in the two images. Therefore, we use a high textured pattern (designed pattern) to compute the matrix (fundamental matrix) representing the relative position between the cameras of

each stereoscopic device. It should be noted that the video cameras are rigidly fixed; hence we can perform the different calibration steps off-line and outside of the patient's abdomen for each camera pair.

3.2   3-D Reconstruction

To reconstruct a 3-D surface of the organs, we use standard stereoscopic reconstruction tools. Using the epipolar geometry principle, the points correspondences are searched between two images (left image and right image) of each stereovision system [2]. Once the calibration steps described in section 2 are performed, we can make the search of the correspondences between the images acquired by the left and right cameras using the epipolar constraint and the Zero Mean Normalized Cross Correlation (ZMNCC) function. For this, a point correspondence search between two images using a multiple resolutions approach and epipolar constraints is used. This allows us to reduce significantly the search area (Figure 6). Moreover, images from both cameras are re-sampled at different resolution levels. For each of them, the reference image is divided into a regular grid. For the lowest resolution level, an initialization step searches a good starting point correspondence among a few points taken from the central region of the image. Once this initial point is found, all the other correspondences are computed using the epipolar constraint and the ZMNCC function. Finally, the result is transmitted as input for the next resolution level.
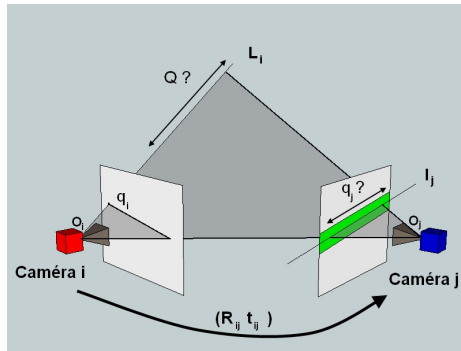

Fig. 6. Illustration of the epipolar constraints

After this 3D points reconstruction; we can build a 3D triangular mesh from the local information (texture) for each 3D point of its close neighborhood. This model of the operating site surfaces can then be used in the whole visualization system to enhance the surgeon's view.

3.3   Color Information

Once we have created a 3D mesh of a part of the abdominal cavity from a stereoscopic pair of cameras, we can en enhance this model with color information.

Our video cameras are not color calibrated. Thus, the colors of a same area may slightly vary. To overcome this issue, we propose to mix the colors so as to create a smooth change of color whilst changing the point of view (*fade in* or *fade out* effect).

There are two ways of giving colors to the 3D vertices of our mesh. Te first solution consists in mapping the corresponding camera model to the mesh. The graphic card then interpolates the whole mesh surface. In this case, the color resolution will be the same as the video camera resolution.

The second solution is to give a color to each point of the 3D mesh. As our mesh is made of triangles, the graphic card can handle the interpolation within each triangle. This latter method allows merging color information from several video cameras on any point of the mesh, however, the color resolution will be the same as the mesh resolution, and so we will have to refine the mesh when it is needed. Then, for each 3-D point P of the 3D mesh, we compute its projection into the space of each camera nc to find its color contribution for this camera $color(p(i,j)_{nc})$ and weight it (α_{nc}) to obtain the final color of P as illustration Equation 1.

$$color(\mathbf{P}) = \sum_{nc=1}^{nc=N} \alpha_{nc} \times color(p(i,j)_{nc}) \qquad (1)$$

The choice of the weights α_{nc} allows to get several visual effects. We indeed can

1. choose the texture of only one camera of one stereoscopic device (all other weights are 0)
2. set all teh coefficients to the same value ($\forall nc \in [1,N], \alpha_{nc} = \frac{1}{N}$) In the case of 2 or more video cameras, this setting generally creates a fuzzy result image.
3. compute α_{nc} as a function of a distance between the camera nc and the virtual camera displaying the 3-D mesh.
4. compute α_{nc} as a function of a distance between the camera nc and the mesh position.

4   Experimental Setup and Results

The distributed vision system, the calibration methods, and the 3-D reconstruction (3-D virtual environment) presented in this paper were validated on an experimental setup simulating a case of laparoscopic surgery (radical prostatectomy) (see Figure 7). Inside the experimental bench representing an abdominal cavity are placed several porcine organs.

4.1   First Benefits of the Developed System

The first benefits of the proposed vision system are: 1) an improved image quality, 2) a high resolution (1600×1200 pixels for each camera), 3) the high quality

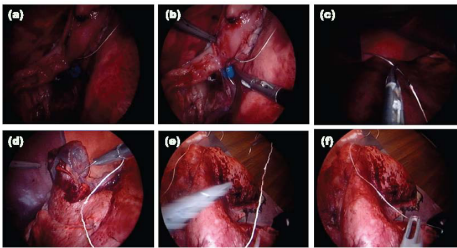**Fig. 7.** The experimental set-up which reproducing a laparoscopic surgery bench

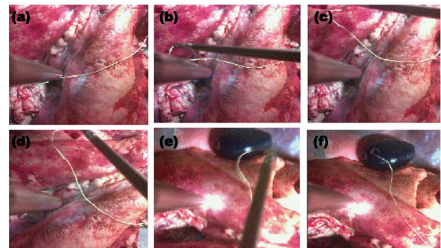**Fig. 8.** Images captured by a conventional endoscope

Fig. 9. Images given by one of our CMOS cameras

lighting that characterizes our vision system compared to the conventional endoscopy vision. Figure 3.(a) and Figure 3.(b) represent the comparison between the two vision systems used at the same time reproducing the conditions of a minimally invasive surgery.

Other advantages of the proposed multiple views system are to give a better FOV (multiplied by $\propto 4$), DOF (multiplied by $\propto 2$) and high resolution images. However, this system can also offer a full and wide virtual 3-D model, high resolution thanks to the presence of multiple cameras with different viewpoints.

In addition, the mechanical solution to suppress specular reflections also directly helps the surgeon to see parts of organs that would have been saturated on the image. Figure 10 illustrates a porcine liver observed with and without our system.
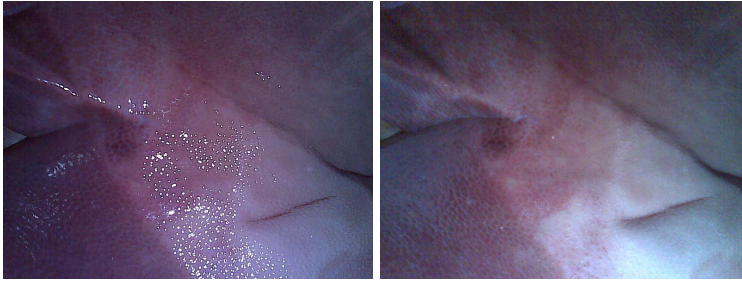
(a) Porcine liver in endoscopic conditions without polarized film

(b) Porcine liver in endoscopic conditions with polarized films

Fig. 10. Illustration of the mechanical suppression of the specular reflections using polarized film

4.2   3-D Virtual Model

To verify our 3-D mesh reconstruction, we made a plaster phantom of a porcine liver (see Figure 11 (a)). The surface was reconstructed from a CT-scanner[2] and with our method and then both surfaces where registered thanks to an ICP[3] algorithm.
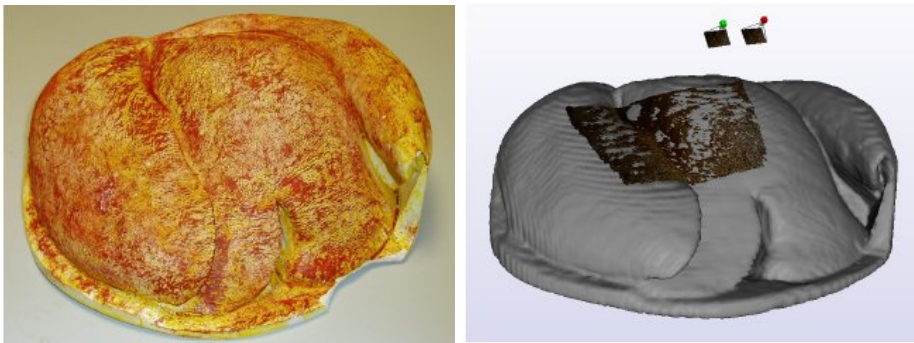
(a) Plaster phantom of porcine liver

(b) Mesh reconstruction applied to the phantom

**Fig. 11.** 3-D mesh reconstruction validation

The errors of the ICP registration is displayed Table 1. Even if the largest error is about $5mm$, more than 90% of points have an absolute error less than $2.1mm$.

---

[2] Computed Tomography-Scanner.
[3] Iterative Closest Point.

Table 1. ICP registration error between CT-reconstructed Mesh and Stereo-vision reconstructed mesh

min	0 mm
max	4.858 mm
mean	0.496 mm
standard deviation	0.323 mm
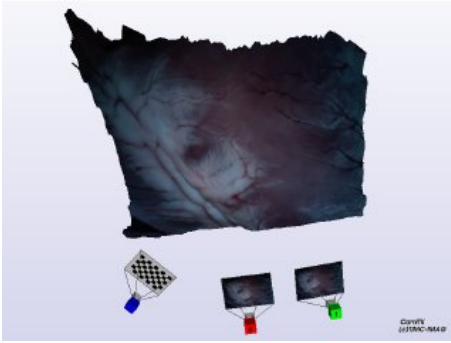
(a) use of only one camera (the red one)
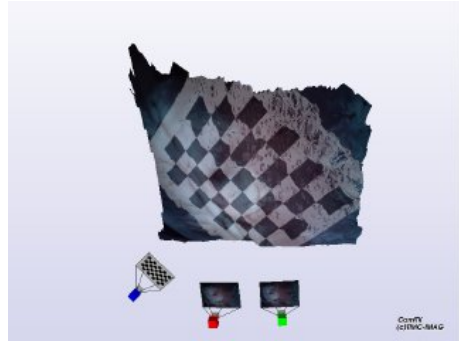
(b) use the same coefficients for the 3 images

(c) use the distance between the real and the virtual camera

(d) use the distance between the camera and the mesh

Fig. 12. Illustration of the 4 possible color strategies. The blue camera has a virtual image for the sake of the illustration.

Concerning the color mapping, we tested the mapping with one real and one virtual image to better see the differences between the 4 different approaches proposed in the previous section. Figure 12 illustrates the obtained results.
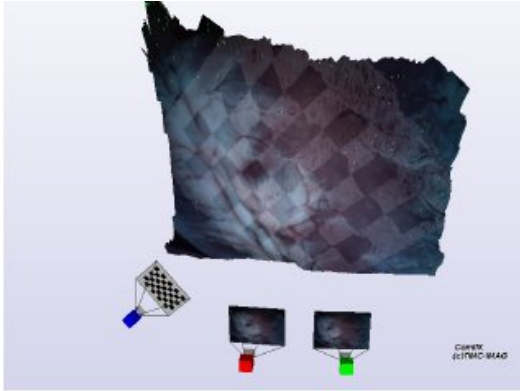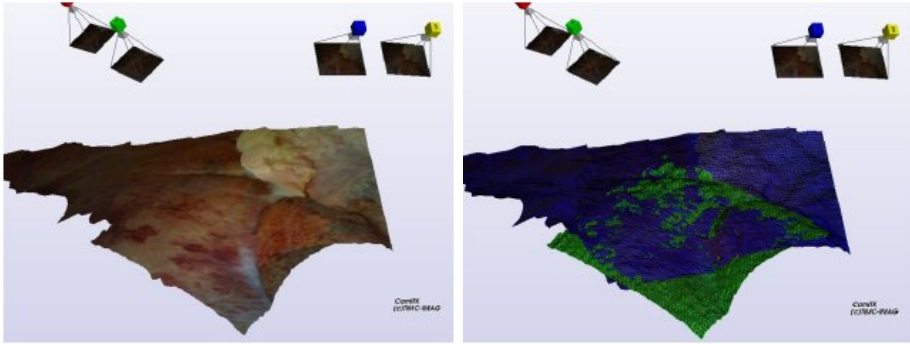
4.3   3D-Mosaicing

To obtain even wider range of view, we can use several stereoscopic devices and build a 3-D mosaic of the reconstructed scene. Right now, the registration between meshes is done manually but further studies will be conducted for an automatic registration (see Figure 13). Improving the 3-D field-of-view (resp. 2-D field-of-view) in the case of laparoscopy surgery allows overcoming the traditional drawbacks of the invisible parts (organs) with a limited field-of-view. It also reduces the communication between the surgeon and his/her assistant for the repositioning of the traditional endoscope (for a better view), which allows the surgeon to focus on his surgical gestures.

(a) Image obtained with 2 stereo-scopic devices

(b) Corresponding mesh

Fig. 13. Example of mesh mosaicing

One of the main objectives of the proposed vision system is to provide the surgeon with any view of the patient's abdominal cavity, without having to physically move his endoscopic camera. It is also possible to zoom in and zoom out on a region of interest while maintaining a high resolution 3-D model thanks to the multiple resolutions 3-D reconstruction technique described in section 3. These techniques are validated on different porcine organs in laparoscopy conditions.

Figure 14 shows examples of a 3-D reconstruction of different porcine organs: heart with or without wound to suture (images (a) and (b)), porcine lung (image (c)) and porcine gall bladder (image (d)). The built model is displayed in a 3-D virtual environment allowing easy manipulation i.e. translation (x and y axes), orientation (pan θ and tilt α), and zoom in/zoom out ($z-$ and $z+$ axis).
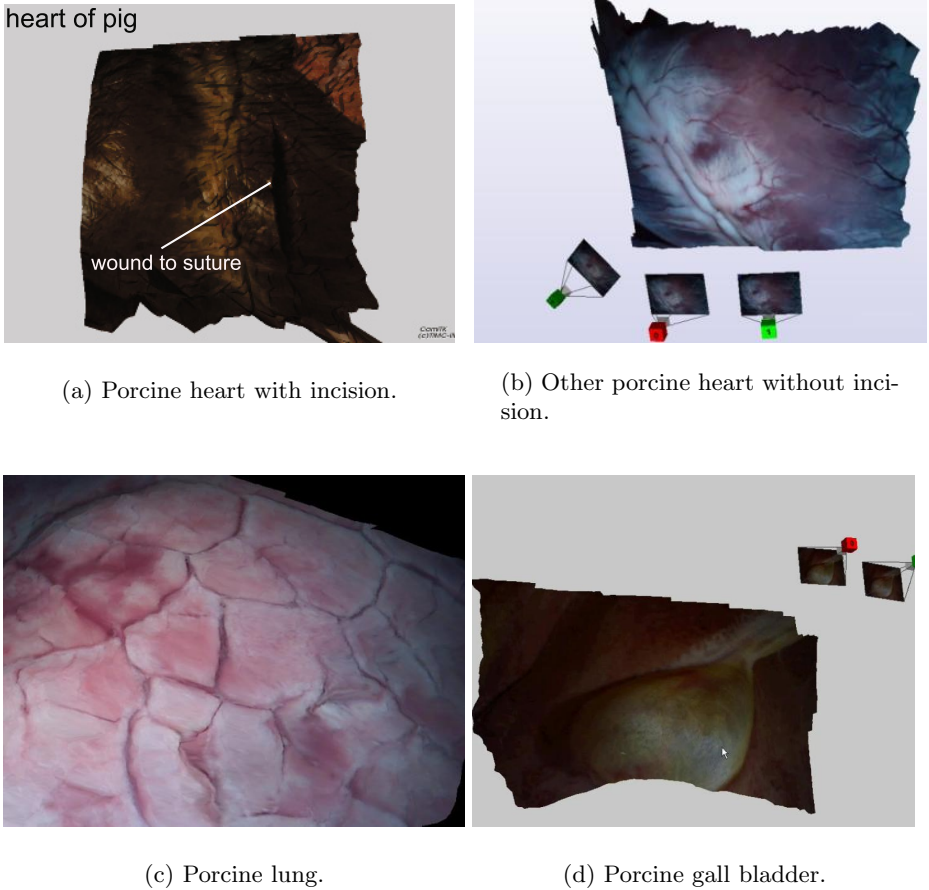
(a) Porcine heart with incision.

(b) Other porcine heart without incision.

(c) Porcine lung.

(d) Porcine gall bladder.

Fig. 14. Illustration of the reconstruction of 3-D model of different porcine organs

5   Discussion

The evaluation of the 3-D reconstruction method presented in section 4 was performed on a plaster phantom whose 3-D model is perfectly known (computed with a CT-scanner). The accuracy of 3-D reconstruction is estimated at less than one millimeter with a computation time of few seconds (5 seconds using a standard PC of 1.8 GHz).

We chose a material approach in order to reduce specular reflections (characterized by bright white stain that can complicate the point-correspondence search for the 3D reconstruction). We still face residual reflections, that can cause problems during the point-matching, and causes a lack of information on the reconstructed 3D model. To solve this issue, a software-based solution could

also be considered : for instance, it would be possible to decompose the image into a specular component and a diffuse component to improve the process.

In order for our approach to be compatible with real-time clinical constraints, we will need to reduce the actual computation time in order to provide the surgeon with a more fluid 3D visualization (3 to 4 3D images / second max). It must be noted that we haven't yet tackled the computation time issue, but the code could be easily optimized and parallelized using GPU programming.

The other aspect of the proposed system that needs to be improved is the necessity to make two extra 5mm incisions to maintain the cameras with lassos (deployment system) continuously during the surgical intervention. We plan on studying other approaches for the deployment system (materials, shape, etc...) allowing for the introduction of all the cameras through an unique incision, and allowing for their stabilization.

It will also be possible to improve the navigation using the virtual camera (i.e. 3-D virtual view): currently, the surgeon interacts with the 3-D reconstruction using a mouse-click but in the future we plan on using more intuitive controls e.g. voice-based navigation, control of the model by performing specific movements with the surgical instruments that could be detected in real-time using image processing approaches [13], etc.

6   Conclusion and Future Works

A new multiple views system was presented in this paper to compensate some of the drawbacks of traditional endoscopy vision. It includes two or four high resolution CMOS camera positioned so as to reproduce stereovision conditions. This vision system solves the problems associated to the use of conventional endoscopic vision as the low field-of-view, the weak depth-of-field, insufficient resolution, and the lack of lighting. In this paper, we have also presented a method to reconstruct a multiple resolutions 3-D virtual environment to help the surgeon to view an organ in three dimensions. This solves the problem of loss of the notion of depth information in the context of indirect vision in minimally invasive laparoscopic surgery. The proposed materials were validated in an experimental setup representing the conditions of laparoscopy surgery (radical prostatectomy) using the developed multiple views system. Future work will involve reducing the computation time required to calculate the 3-D virtual models, currently estimated at 5 seconds. To do, a part of the 3-D reconstruction algorithm (especially the matching step) may be transferred to a GPU[4] computing. After these successful initial validation tests, we plan to continue these tests on human cadavers.

---

[4] Graphics Processing Unit.

References

1. Baumhauer, M., Simpfendoerfer, T., Schwarz, R., Seitel, M., Muller-Stich, B.P., Gutt, C.N., Rassweiler, J., Meinzer, H.-P., Wolf, I.: Soft tissue navigation for laparoscopic prostatectomy: Evaluation of camera pose estimation for enhanced visualization. In: Medical Imaging 2007: Visualization and Image-Guided Procedures, vol. 6509(1) (2007)
2. Faugeras, O.: Three-dimensional computer vision: a geometric viewpoint. MIT Press, Cambridge (1993)
3. Feuerstein, M., Mussack, T., Heinning, S.M., Navab, N.: Registration-free laparoscope augmentation for intra-operative liver resection planning. Progress in Biomedical Optics and Imaging 8(1), 28 (2007)
4. Fowler, D.L., Hu, T., Nadkarni, T., Allen, P.K., Hogle, N.J.: Initial trial of a stereoscopic, insertable, remotely controlled camera for minimal access surgery. Surgical Endoscopy 24(1), 9–15 (2010)
5. Harris, C., Stephens, M.: A combined corner and edge detector. In: Alvey Conference, Manchester, pp. 147–151 (1988)
6. Karagozler, M.E., Cheung, E., Kwon, J., Sitti, M.: Miniature endoscopic capsule robot using biomimetic micro-patterned adhesives. In: The First IEEE/RAS-EMBS Int. Conf. on Biomedical Robotics and Biomechatronics, pp. 105–111 (February 2006)
7. Lerotic, M., Chung, A.J., Mylonas, G.P., Yang, G.Z.: pq-space based non-photorealistic rendering for augmented reality. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 102–109. Springer, Heidelberg (2007)
8. Lin, S., Li, Y., Kang, S.B., Tong, X., Shum, H.-Y.: Diffuse-specular separation and depth recovery from image sequences. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part III. LNCS, vol. 2352, pp. 210–224. Springer, Heidelberg (2002)
9. Makhoul, B., De La Taille, A., Vordos, D., Salomon, L., Sebe, P., Audet, J., Ruiz, L., Hoznek, A., Antiphon, P., Cicco, A., Yiou, R., Chopin, D., Abbou, C.: Laparoscopic radical nephrectomy for t1 renal cancer: The gold standard? A comparison of laparoscopic vs open nephrectomy. BJU Int. 93(1), 67–70 (2004)
10. Mylonas, G.P., Kwok, K.-W., Darzi, A., Yang, G.-Z.: Gaze-contingent motor channelling and haptic constraints for minimally invasive robotic surgery. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 676–683. Springer, Heidelberg (2008)
11. Robb, R.A., Cameron, B.M., Aharon, S.: Efficient shape-based algorithms for modeling patient specific anatomy from 3d medical images: applications in virtual endoscopy and surgery. In: Proceedings of the 1997 International Conference on Shape Modeling and Applications, pp. 97–108, 245–246 (March 1997)
12. Salvi, J., Pagès, J., Batlle, J.: Pattern codification strategies in structured light systems. Pattern Recognition 37(4), 827–849 (2004)

13. Wolf, R., Duchateau, J., Cinquin, P., Voros, S.: 3D tracking of laparoscopic instruments using statistical and geometric modeling. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part I. LNCS, vol. 6891, pp. 203–210. Springer, Heidelberg (2011)
14. Wu, C.-H., Sun, Y.-N., Chang, C.-C.: Three-dimensional modeling from endoscopic video using geometric constraints via feature positioning. IEEE Transactions on Biomedical Engineering 54(7), 1199–1211 (2007)
15. Xie, T., Guo, S., Chen, Z., Mukai, D., Brenner, M.: Grin lens rod based probe for endoscopic spectral domain optical coherence tomography with fast dynamic focus tracking. Optics Express 14, 3238–3246 (2006)
16. Zhang, Z.: A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(11), 1330–1334 (2000)

Generation of Synthetic 4D Cardiac CT Images by Deformation from Cardiac Ultrasound

Feng Li[1,2], James A. White[1,3], Martin Rajchl[1,2],
Aashish Goela[4], and Terry M. Peters[1,2]

[1] Imaging Research Laboratories, Robarts Research Institute, London, ON
[2] Biomedical Engineering Graduate Program, Western University, London, ON
[3] Division of Cardiology, Department of Medicine, Western University, London, ON
[4] Department of Medical Imaging, Western University, London, ON

Abstract. In this paper we introduce a method of generating a synthetic 4D cardiac CT dataset using a single (static) CT and 4D ultrasound images. The method performs non-rigid registrations among ultrasound images to obtain deformation fields of patient specific heart motion and applies the deformation fields onto a static CT image to deform it into a series of dynamic CT images. It is the ultimate aim that this novel synthetic CT dataset could be used to generate high spatiotemporal resolution 4D cardiac models for the intra-operative guidance of minimally invasive cardiac intervention. Validations were performed by comparing synthetic CT images to real dynamic CT images.

Keywords: image guided intervention, augmented reality, minimally invasive cardiac intervention, synthetic CT, non-rigid registration.

1   Introduction

Conventional cardiac surgery procedures have applied a mid-line sternotomy to access the patient's heart. However, minimally invasive approaches that limit the need for thoracic trauma have rapidly emerged as clinically preferred techniques for many procedures, and may also negate the requirement for arresting the heart (i.e.: beating heart surgery) [1, 2]. However, by their very nature, these procedures lack direct visualization of surgical targets and/or tools, a challenge that is compounded by the potential of the target movements within the cardiac cycle. Therefore sophisticated image-guided navigation systems are required to assist in procedural efficiency and therapeutic success.

For minimally invasive beating heart surgical procedures, the optimal navigation system would show both the pertinent cardiac anatomy and dynamic motion of the surgical targets throughout the cardiac cycle. It should also be cost efficient and be easily integrated into standard workflow within the operation room. Such a goal is difficult to successfully accomplish using a single imaging modality. For example, endovascular visualization is not appropriate as blood will blocks the view of target structures, and while intra-operative MRI

has the capability of dynamically displaying cardiac structures and their motion during interventions [3], this technique is very expensive and technically challenging with the need to develop novel, non-ferromagnetic tools and devices. Intra-operative trans-esophageal ultrasound is more accessible and clinically feasible. However, while substantial improvements in 3D acquisition protocols have occurred, the field of view remains relatively restricted and limits modeling of relevant cardiac and vascular structures beyond this region. Further, artifacts, such as acoustic shadowing, can limit the capacity to accurately develop volumetric models. Another cardiac imaging technique, cardiac gated CT, has emerged as a superior tool for anatomic imaging of cardiac structures through isotropic voxel sizes of approximately 0.5mm [17]. While retrospective gating can be applied to obtain dynamic volumetric imaging, the temporal resolution is limited to gantry rotation speed and radiation dose exposure is at least four times higher than for a single volume [7]. Accordingly, no single imaging test appears to be ideal for the provision of 4D cardiac modeling to plan and guide cardiovascular procedures.

We propose a solution that would allow a high-resolution 4D cardiac model to be generated from a synthetic CT dataset derived from a single pre-operatively acquired 3D CT that has been iteratively deformed to create a dynamic 4D sequence using 4D ultrasound data. These pre-operatively generated images would be used to provide a high quality 4D anatomical context through visualization of some or all structural components spatially co-registered (and ECG-synchronized) to the patient using standard intra-operative imaging, such as 2D/3D ultrasound or fluoroscopy. With these real-time registration techniques, an augmented reality framework can be developed and integrated into navigation platforms to provide critical information to surgeons at the time of intervention.

In our previous work, we proposed using pre-operative high spatial resolution CT and intra-operative 4D ultrasound images within a navigation system[4, 5]. Both of these tests are commonly used in standard clinical workflow. However, prior experiments have shown that registering a 4D ultrasound sequence to a single, static CT image will result in large target registration errors (TRE) at non-corresponding cardiac phases because of cardiac deformation [8]. In the context of a dynamic CT dataset, a rapid registration method has been shown to offer co-registration of corresponding CT frames to cardiac ultrasound frames based upon ECG gating information [6] and may offer improved TRE values.

Much work has been performed towards estimating cardiac motion in ultrasound and MRI images by using non-rigid registration [9-12], with most of the reports suggesting that the estimated motion can be used for the assessment of cardiac motion and mechanics. In this paper, we introduce a new concept that attempts to combine the motion estimation obtained from dynamic ultrasound images with a static CT image to generate a synthetic dynamic CT dataset. These synthetic CT images can take advantage of the higher temporal resolution of the ultrasound images upon which they are modeled. By introducing these synthetic CT images into a navigation system, we anticipate improved model to patient registration and more intuitive visualization of therapeutic targets for beating heart surgeries.

Validation experiments were performed by comparing the synthetic CT images to ground truth dynamic CT images. Dice similarity coefficients (DSC) and RMS errors were computed for the left ventricle.

2   Method

2.1   Overall Workflow

The suggested workflow for generating synthetic CT images and integrating the images into clinical workflow is described below (Fig 1). A high spatial resolution static CT image and a sequence of high temporal resolution dynamic ultrasound images are acquired from the patient pre-operatively. Usually, static cardiac CT images are obtained from prospectively gated CT scans and represent the mid-diastole phase. Dynamic ultrasound images in this project are obtained by using a real-time transesophageal echocardiogram (TEE) and represent at least one complete cardiac cycle. Non-rigid registrations are first performed on the dynamic ultrasound images to obtain deformation fields representing heart motion maps. Deformation fields are then applied to the static CT image to deform it into a series of synthetic dynamic CT images. Anatomical features can be defined and models can be developed based on the synthetic images as required. The synthetic CT images and models are then transferred to the navigation system peri-operatively. During the operation, intra-operative ultrasound images, transferred from the ultrasound machine to the navigation system, are continuously registered to the synthetic CT images at corresponding cardiac phases, based on gating information. Surgeons can choose different visualization modes that provide different fusion of ultrasound image, CT images, and models.

2.2   Generating Synthetic CT Images

The methodology to generate synthetic dynamic CT images is to perform non-rigid registrations between ultrasound images within a single cardiac cycle to obtain patient specific heart motion maps, in the form of deformation fields, and to apply these vector maps to CT images to provide synthetic animation.(Fig. 2) In this approach, at least one sequence of 3D TEE images, representing at least one complete cardiac cycle, and one single frame cardiac CT images must be acquired pre-operatively.

The procedure begins with selecting one ultrasound image, which has the same or closest cardiac phase as the static CT image, as a reference. A rigid registration [8] is then performed between this image and the static CT image. The rigid registration begins with semi-automatically segmenting the inner wall of the left ventricle [18] from both the static CT image and the reference ultrasound image and use the iterative closest point (ICP) method to align the surfaces. The alignment is then refined by a mutual information based registration [16] resulting in an optimized transform. The transform is then applied on the reference image and all the other ultrasound images in the 4D sequence to transform them into the CT coordinate system.
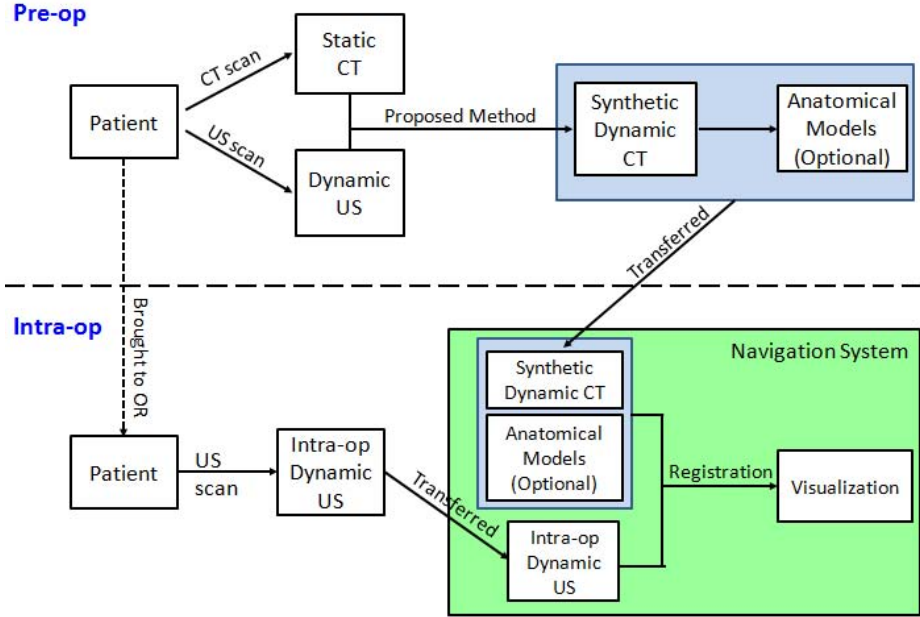
Fig. 1. Overall workflow showing how synthetic dynamic CT images are generated and integrated into a navigation system

After the initial rigid registration step, non-rigid registrations are performed among the ultrasound images in the 4D sequence and deformation fields are recorded. This can be achieved by either registering the reference image directly to every image in the sequence, or performing a registration between adjacent frames. The deformation fields obtained from the non-rigid registration are used as heart motion maps and applied to the static CT image to generate a synthetic dynamic CT sequence. By performing the approach for each frame, we generate an entire sequence of synthetic dynamic CT images with the same temporal resolution as the dynamic ultrasound images. The non-rigid registration method that we use in this paper is a multi-resolution fast free-form (F3D) deformation registration method [13], while several non-rigid registration methods, such as Demons algorithm [14] and Level Set Motion method [15] will be used and compared in later work.

3   Experiments and Results

In the experiments, we used dynamic CT images as the gold standard. The dynamic CT images were obtained from retrospectively gated CT scans and reconstructed as ten frames per cardiac cycle. The first frame is at mid-diastole. Patient-specific pre-operative TEE images were obtained as well. The first CT frame was chosen as the image from which the synthetic dynamic image set
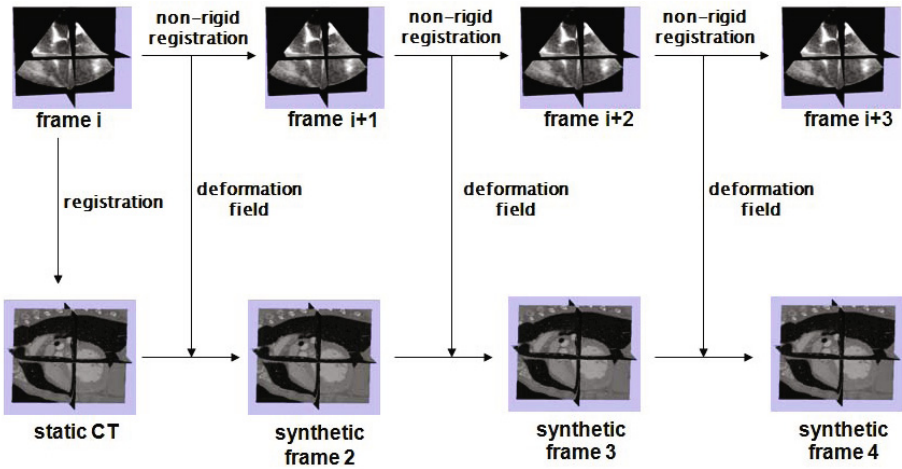
Fig. 2. Demonstration for how to generate synthetic CT images

was constructed. Figure 3 shows two slices from the synthetic images, one at end-diastole and the other at end-systole.

Then we compared the synthetic CT images to the original dynamic CT images. Since synthetic and original CT images have different frame rates, for each original CT image, we manually select the synthetic CT image with the closest cardiac phase for comparison. The ventricular blood pools were segmented from both the synthetic and original images. Dice similarity coefficients (DSC) of the blood pool volumes and mean RMSE of the ventricle surfaces were computed between each pair. The results are listed in Table 1.
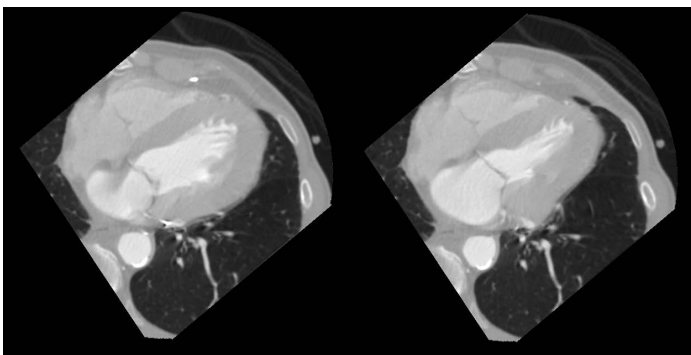

Fig. 3. Synthetic images. Left: a slice at end-diastole frame Right: a slice at end-systole frame.

Table 1. Comparison between synthetic and original CT images

# of frames	1	2	3	4	5	6	7	8	9	10	mean
Dice metric	0.86	0.86	0.78	0.59	0.59	0.80	0.79	0.75	0.78	0.83	0.76
RMS (mm)	1.54	1.70	2.23	4.46	3.85	2.13	1.96	2.47	2.44	2.04	2.48

Table 2. Comparison between synthetic and original CT images with modifications on papillary muscles

# of frames	1	2	3	4	5	6	7	8	9	10	mean
Dice metric	0.86	0.85	0.78	0.69	0.68	0.79	0.79	0.74	0.78	0.83	0.78
RMS (mm)	1.54	1.83	2.35	3.44	3.50	2.13	1.96	2.47	2.44	2.04	2.37

It is noticeable that frame 4 and 5, which are very close to end-systole, have relatively lower dice metrics and higher RMS values. By visualizing and comparing the surfaces from frame 5 (Fig. 4b), we can see the volume of blood pool from the synthetic image is larger than that from original data. One reason for this difference is likely to be the difference in imaging papillary muscles in CT and ultrasound. Papillary muscles become much thicker at end-systole and this feature can be seen clearly in CT images. However, ultrasound images do not show papillary muscles clearly, so the deformation fields obtained from ultrasound images in this region may not represent the change of papillary muscles very well, and cause the non-rigid registrations to under-estimate the deformation of the left ventricle.

To estimate the impact of papillary muscles, we modified the segmentations from original CT images to include papillary muscles and re-compute the dice metrics and RMS values, (Table 2), where the DSC of frame 4 and frame 5 were increase by about 10% and the RMSE values were decreased by 1.02 mm and 0.35 mm respectively.

We also computed short-axis slicewise DSC from basal to apical to obtain an error map along the longitudinal axis. For most of the frames, the slicewise DSC changed slightly along the longitudinal axis. However, in frame 4 and 5, the DSC remained high around mitral valve area, while it decreased significantly when approaching the apex. (Fig. 5)

4   Discussion

The validation based on DSC and RMSE showed good results for most of the frames in a single cardiac phase, while there were some issues with two systole frames. In the slice-by-slice DSC diagram (Fig. 5), we can see that the DSC remained high at the upper half of the ventricle for all the frames. However, two systole frames presented low DSC at the lower half of the ventricle. Looking into those slices with low DSC, it is evident that the surfaces from the synthetic CT images completely contained the surfaces from the original CT images, which implies that the synthetic CT images under-estimate the ventricle contraction.
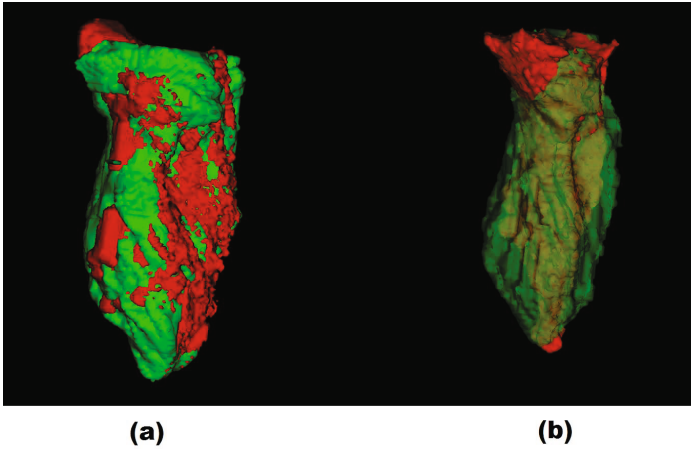
Fig. 4. Comparison between synthetic and original CT images. Left: surfaces of inner wall of left ventricle at end-diastole. Right: surfaces of inner wall of left ventricle at end-systole (original: red, synthetic: green)

By visualizing the CT and ultrasound image together, we observed that this issue was mainly caused by the difference in the left ventricle representation between CT and ultrasound images. Figure 6 shows a pair of registered CT and ultrasound volumes. The CT frame shows frame 5 from the original dynamic CT series, while the ultrasound image has the closest cardiac phase as the CT images. The outline of the left ventricles from ultrasound and CT images were manually delineated as red and yellow curves respectively. Clearly, the red curve completely contains the yellow curve, conferring what we observe with the DSC results.

Thickened papillary muscles are clearly visualized in the CT, but not ultrasound images, resulting in a significant difference between the two images. Temporal registration issues can also cause inconsistency, since the ultrasound images in the two data sets have very low frame rate, i.e. 7 frames per cardiac cycle[1] , much lower than common ultrasound images. The low frame rate makes it difficult to align the ultrasound and CT images temporally and a small misalignment may lead to large difference in the shape of left ventricle because of the fast myocardial motion at systole. The third reason for inconsistent results between the image datasets is related to the heart motion during the CT and ultrasound scans. In CT scans, β-blockers are commonly used to reduce patients' heart rate, while this does not apply for ultrasound scans. The usage of β-blockers may change the ventricular contraction leading to different heart motion in CT and ultrasound images. However, this may promote the motivation

---

[1] The most common frame rate of images from the same ultrasound machine is about 25 frames per second, and it also has the capability of imaging more than 30 frames per second.
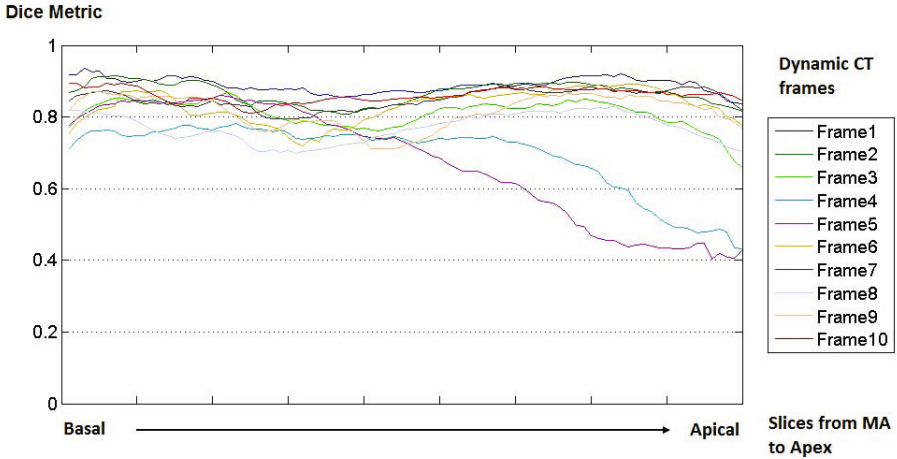
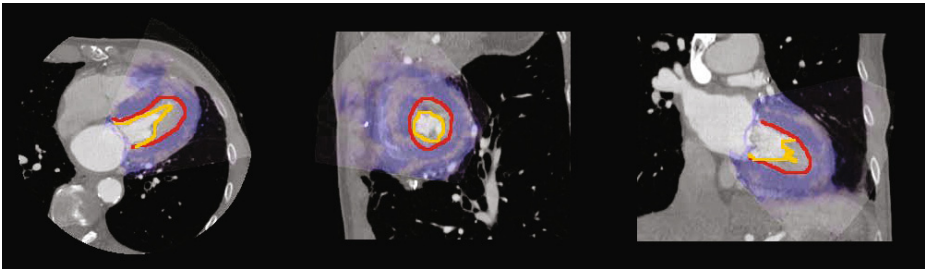**Fig. 5.** Slice-by-slice dice metrics along the longitudinal axis

Fig. 6. Comparison of the left ventricles from CT and ultrasound images at systole. Ultrasound images are shown in blue. The outline of the left ventricle from the ultrasound image is manually indicated as a red curve, while the outline from CT is indicated as a yellow curve.

to employ synthetic CT images, because the motion represented in the synthetic images is derived from pre-operative ultrasound, which should represent similar heart motion to intra-operative ultrasound images.

Despite some potential limitations discussed above, this method has many advantages. For example, it is much more cost efficient than using intra-operative MRI and does not require non-ferromagnetic tools or devices. It can be very easily integrated into current clinical workflow and the OR, since TEE and CT are already used routinely and frequently in standard workflow. Compared to retrospectively gated CT scans, this method applies much less radiation dose to patients. It takes advantage of both ultrasound's high temporal resolution and CT's high spatial resolution, while none of an existing imaging modality can provide both of them. It's patient-specific and aims to deal with each patient's

individual cardiac motion, patients may have abnormal motion patterns, rather than relying on an average model. In the guidance system, it can lead to better registration results than using static CT image and can be potentially visualized in different ways according to user's requirement, for example through volume rendering which does not apply to surface models.

5   Conclusion and Future Work

This paper introduced a novel methodology that attempts to exploit the high temporal resolution of ultrasound imaging and high spatial resolution of CT imaging to generate a novel, high spatio-temporal resolution synthetic CT. Our initial results show the feasibility of this method and sets the foundation for future work validating its accuracy for modeling anatomic cardiac motion in different disease states and validating the implementation of cardiac model development for the guidance of minimally invasive procedures within augmented reality environments.

References

1. Dotty, D.B., Flores, J.H., Doty, J.R.: Cardiac valve operations using a partial sternotomy technique. J. Card. Surg. 15, 35–42 (2000)
2. Vassiliades, T.A., Block, P.C., Cohn, L.H.: The clinical development of percutaneous heart valve technology. J. Thorac. Cardiovasc. Surg. 129, 970–976 (2005)
3. McVeigh, E.R., Guttman, M.A., Kellman, P., Raval, A.A., Lederman, R.J.: Real-time interactive MRI for cardiovascular interventions. Acad. Radiol. 12, 1221–1227 (2005)
4. Linte, C.A., Moore, J., Wedlake, C., Bainbridge, D., Guiraudon, G.M., Jones, D.L., Peters, T.M.: Inside the beating heart: An in vivo feasibility study on fusing pre- and intra-operative imaging for minimally invasive therapy. Journal of Computer Assisted Radiology and Surgery 4(2), 113–123 (2009)
5. Linte, C.A., Moore, J., Wiles, A.D., Wedlake, C., Peters, T.M.: Virtual reality-enhanced ultrasound guidance: A novel technique for intracardiac interventions. Comput. Aided Surg. 13(2), 82–94 (2008)
6. Huang, X., Hill, N.A., Ren, J., Guiraudon, G., Boughner, D., Peters, T.M.: Dynamic 3D ultrasound and MR image registration of the beating heart. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 171–178. Springer, Heidelberg (2005)
7. Shuman, W.P., Branch, K.R., May, J.M., Mitsumori, L.M., Lockhart, D.W., Dubinsky, T.J., Warren, B.H., Caldwell, J.H.: Prospective versus retrospective ECG gating for 64-detector CT of the coronary arteries: comparison of image quality and patient radiation dose. Radiology 248(2), 431–437 (2008)
8. Li, F., Lang, P., Rajchl, M., Chen, E.C.S., Guiraudon, G., Peters, T.M.: Towards real-time 3D US-CT registration on the beating heart for guidance of minimally invasive cardiac interventions. In: Proc. SPIE, vol. 8316, p. 831615 (2012)
9. Ledesma-Carbayo, M.J., Kybic, J., Desco, M., Santos, A., Suhling, M., Hunziker, P., Unser, M.: Spatio-temporal nonrigid registration for ultrasound cardiac motion estimation. IEEE Trans. Med. Imaging 24(9), 1113–1126 (2005)

10. Shi, W., Zhuang, X., Wang, H., Duckett, S., Luong, D.V., Tobon-Gomez, C., Tung, K., Edwards, P.J., Rhode, K.S., Razavi, R.S., Ourselin, S., Rueckert, D.: A Comprehensive Cardiac Motion Estimation Framework Using Both Untagged and 3-D Tagged MR Images Based on Nonrigid Registration. IEEE Trans. Med. Imaging 31(6), 1263–1275 (2012) (Epub February 15, 2012)
11. Wierzbicki, M., Drangova, M., Guiraudon, G.M., Peters, T.M.: Validation of dynamic heart models obtained using non-linear registration for virtual reality training, planning, and guidance of minimally invasive cardiac surgeries. Medical Image Analysis 8(3), 387–401 (2004)
12. Sundar, H., Litt, H., Shen, D.: Estimating myocardial motion by 4D image warping. Journal Pattern Recognition 42(11), 2514–2526 (2009)
13. Modat, M., Taylor, Z.A., Barnes, J., Hawkes, D.J., Fox, N.C., Ourselin, S.: Fast free-form deformation using graphics processing units. Comput. Meth. Prog. Bio. 98(3), 278–284 (2010)
14. Peyrat, J.-M., Delingette, H., Sermesant, M., Pennec, X., Xu, C., Ayache, N.: Registration of 4D time-series of cardiac images with multichannel Diffeomorphic Demons. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 972–979. Springer, Heidelberg (2008)
15. Vemuri, B.C., Ye, J., Chen, Y., Leonard, C.M.: Image registration via level-set motion: applications to atlas-based segmentation. Medical Image Analysis 7, 1–20 (2003)
16. Pluim, J.P.W., Antoine Maintz, J.B., Viergever, M.A.: Mutual information based registration of medical images: a survey. IEEE Trans. Med. Imaging 22(8), 986–1004 (2003)
17. Rogalla, P., Kloeters, C., Hein, P.A.: CT technology overview: 64-slice and beyond. Radiol. Clin. North Am. 47(1), 1–11 (2009)
18. Rajchl, M., Yuan, J., Ukwatta, E., Peters, T.M.: Fast interactive multi-region cardiac segmentation with linearly ordered labels. In: 9th IEEE International Symposium on Biomedical Imaging (ISBI), pp. 1409–1412 (2012)

Interactive OCT Annotation and Visualization for Vitreoretinal Surgery

Marcin Balicki[1], Rogerio Richa[1], Balazs Vagvolgyi[1], Peter Kazanzides[1]
Peter Gehlbach[2], James Handa[2], Jin Kang[1], and Russell Taylor[1]

[1] ERC for Computer Integrated Surgery, Johns Hopkins University
[2] Wilmer Eye Institute, Johns Hopkins Medical, Baltimore, MD, USA
{marcin,rht}@jhu.edu

Abstract. Vitreoretinal surgery is an extremely challenging surgical discipline requiring surgeons to use limited visualization to locate and safely operate on particularly delicate eye structures. Intraocular image guidance can potentially aid in localizing retinal targets, such as Epiretinal Membranes. This paper describes a system and methods for localizing difficult to identify anatomical features in the retina using video stereo-microscopy and intraocular OCT. We visually track the retina motion and relative position of a hand-held OCT probe to assemble an *M-Scan*, the cross-sectional image of the anatomy corresponding to a trajectory of the probe across the retina. The surgeon is then able to interrogate the OCT image during the procedure by pointing a surgical instrument at the M-Scan trajectory superimposed on the retina and displayed in 3D. The system is designed to provide relevant intraoperative imaging to increase surgical precision, and minimize the surgeon's cognitive load. We describe our system and quantify its performance in a phantom eye.

Keywords: vitreoretinal surgery, image guided interventions, stereo microscopy, anatomical annotations, visualization, OCT, smart instruments.

1 Introduction

Vitreoretinal surgery treats many sight-threatening conditions, the incidences of which are increasing due to the diabetes epidemic and an aging population. It is one of the most challenging surgical disciplines due to its inherent micro-scale, and to many technical and human physiological limitations such as intraocular con-

Fig. 1. A) Image of ERM peeling and B) corresponding OCT B-Scan of the retina [2]

straints, poor visualization, hand tremor, lack of force sensing, and surgeon fatigue. Epiretinal Membrane (ERM) is a common condition where 10 - 80 µm thick scar tissue grows over the retina and causes blurred or distorted vision [1]. Surgical removal of an ERM involves identifying or creating an "edge" that is then grasped and
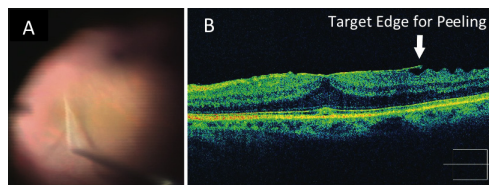
peeled. In a typical procedure, the surgeon uses a stereo-microscope, a vitrectomy system and an intraocular light guide to completely remove the vitreous from the eye to access to the retina. Then, to locate the transparent ERM and identify a potential target edge, the surgeon relies on a combination of pre-operative fundus and Optical Coherence Tomography (OCT) images, direct visualization often enhanced by coloring dyes, as well as mechanical perturbation in a trial-and-error technique [2]. Once an edge is located, various tools can be employed, such as forceps or a pick, to engage and delaminate the membrane from the retina while avoiding damage to the retina itself. It is imperative that all of the ERM is removed, which can be millimeters in diameter, often requiring a number of peels in a single procedure.

The localization of the candidate peeling edges is difficult. Surgeons rely on inconsistent and inadequate preoperative imaging due to developing pathology, visual occlusion, and tissue swelling and other direct effects of the surgical intervention. Furthermore, precision membrane peeling is performed under very high magnification, visualizing only a small area of the retina (~5-15%) at any one time. This requires the surgeon to mentally register sparse visual anatomical landmarks with information from pre-operative images, and also consider any changes in retinal architecture due to the operation itself.

To address this problem we developed a system for intraoperative imaging of retinal anatomy. It combines intraocular OCT with video microscopy and an intuitive visualization interface to allow a vitreoretinal surgeon to directly image sections of the retina using a single-fiber OCT probe and then to inspect these tomographic scans interactively, at any time, using a surgical tool as a pointer. The location of these "M-Scans" is registered and superimposed on a 3D view of the retina. We demonstrate how this system is used in a simulated ERM imaging and navigation task.
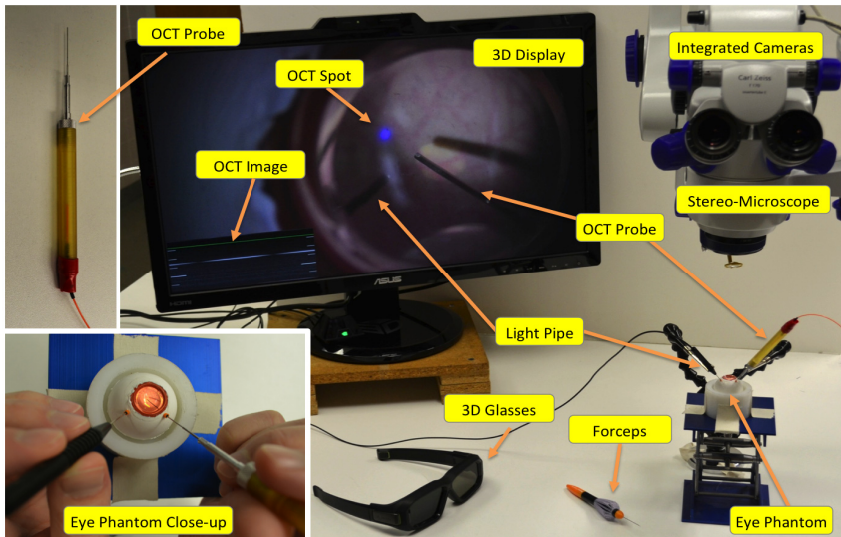

Fig. 2. Components of the Imaging and Visualization System

An alternative approach involves the use of a surgical microscope with integrated volumetric OCT imaging capability such as the one built by Ehlers et al [3]. Their system is prohibitively slow; requires ideal optical quality of the cornea and lens; and lacks a unified display, requiring the surgeon to look away from the surgical field to examine the OCT image, increasing the risk of inadvertent collision between tools and delicate inner eye structures. Fleming et al proposed registering preoperative OCT annotated fundus images with intraoperative microscope images to aid in identifying ERM edges [4], however, they did not present a method to easily inspect the OCT information during a surgical task. It is also unclear whether preoperative images would prove useful if the interval between the preoperative image acquisition and surgery permits advancement of the ERM. Other relevant work uses OCT scanning probes capable of real-time volumetric images [5], but these are still too large and impractical for clinical applications. A single fiber OCT probe presented in [6] has a practical profile but the associated system does not provide any visual navigation capability. Conceptually close works in other medical domains [7-9] have not been applied to retinal surgery, and all, except for [9], rely on computational stereo that is very difficult to achieve in vitreoretinal surgery due to the complicated optical path, narrow depth of field, extreme image distortions, and complex illumination conditions.

2   System Overview

At the center of the system is a visualization system that captures stereo video from the microscope, performs image enhancement, retina and tool tracking, manages annotations, and displays the results on a 3D display. The surgeon uses the video display along with standard surgical tools, such as forceps and a light pipe, to maneuver inside the eye. The OCT image data is acquired with a handheld probe and sent to the visualization workstation via Ethernet. Both applications are developed using *cisst-saw* open-source c++ framework [10] for its stereo-vision processing, multithreading, and inter-process communication. Data synchronization between machines relies on Network Time Protocol.

With the above components, we have developed an imaging and annotation functionality called M-Scan, which allows a surgeon to create a cross-sectional OCT image of the anatomy and review it using a single visualization system. For example, the surgeon inserts the OCT probe into the eye, through a trocar, so that the tip of the instrument is positioned close to the retina and provides sufficient tissue imaging depth. The surgeon presses a foot pedal while translating the probe across a region of interest. Concurrently, the system is tracking the trajectory of the OCT relative to the retina in the video and recording the OCT data, as illustrated in Fig. 3A. The surgeon can add additional M-Scans by repeating the same maneuver. The location of these M-Scans is internally annotated on a global retina map and then projected on the current view of the retina. The surgeon reviews the scan by pointing a tool at a spot on the M-Scan trajectory while the corresponding high- resolution section of the OCT image is displayed, see Fig. 3B.
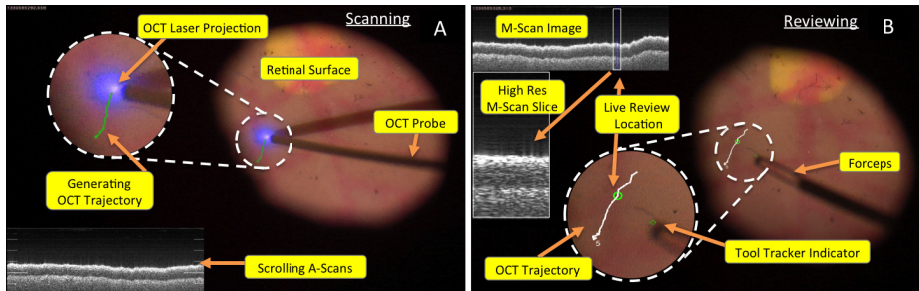
Fig. 3. User Interface: A) Creating M-Scan with OCT probe. B) Review mode with forceps as input.

2.1   Optical Coherence Tomography

OCT is a popular, micron-resolution imaging modality that can be used to image the cross-section of the retina to visualize ERMs, which appear as thin, highly reflective bands anterior to the retina. We developed a common path Fourier domain OCT sub-system described fully in [11]. It includes an 840nm laser source (SLED) with a spectral width of 50nm. A custom built spectrometer is tuned to provide a theoretical axial resolution of 6.2μm (at FWHM) and a practical imaging range of ~2mm in water when used with single fiber probes. The OCT probes are made using standard single mode fiber, with 9μm core, 125μm cladding, and 245μm dia. outer coating, bonded inside a 25 Ga. hypodermic needle. Although, OCT imaging can be incorporated into other surgical instruments such as hooks [6] and forceps, we chose a basic OCT probe because this additional functionality is not required for the experiments where actual peeling is not performed. The system generates continuous axial scan images (A-Scan is 1x1024 pixels) at ~4.5 kHz with latency less than 1ms. The imaging width of each A-Scan is approximately 20-30μm at 0.5-1.5mm imaging depth [12]. The scan integration time is set to 50μs to minimize motion artifacts but is high enough to produce highly contrasting OCT images. By moving a tracked probe laterally, a sample 2D cross-sectional image can be generated. The OCT images are built and processed locally and sent along with A-Scan acquisition timestamps to the visualization station.

2.2   Visualization System

The visualization system uses an OPMI Lumera 700 (Carl Zeiss Meditec) operating stereo-microscope with two custom built-in, full-HD, progressive cameras (60hz at 1920x1080px resolution). The cameras are aligned mechanically to have zero vertical disparity. The 3D progressive LCD display is 27" with 1920x1080px resolution (Asus VG278) and is used with active 3D shutter glasses worn by the viewer. The visualization application has a branched video pipeline architecture [11] and runs at 20-30fps on a multithreaded PC.   It is responsible for stereo video display and archiving; annotation logic; and the retina and tool tracking described below. The following algorithms operate on an automatically segmented square region of interest (ROI)

centered on the visible section of the retina. For the purpose of prototyping the M-Scan concept, this small section of the retina is considered planar for high magnifications. The tracking results are stored in a central transformation manager used by the annotation logic to display the M-Scan and tool locations.
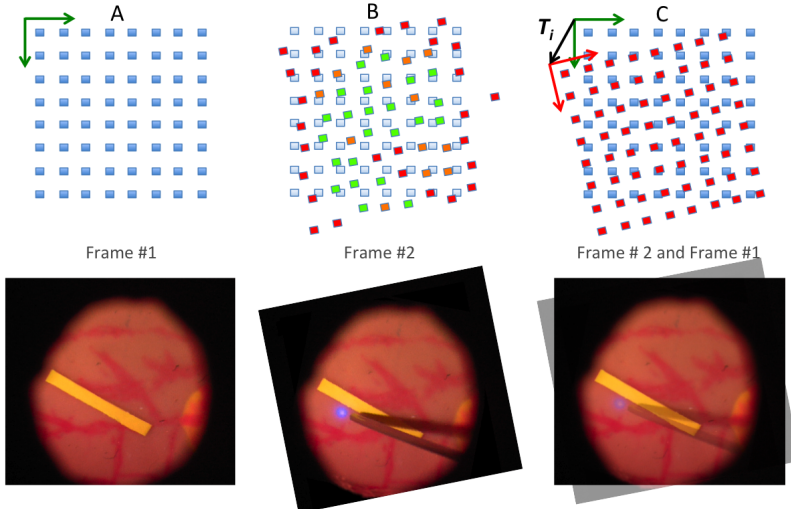

Fig. 4. A) Structured template grid of retina model. B) Templates matching with candidate image. Colors show level of match confidence: red is low, orange is medium, and green is high. Note: matching confidence is low over the tool and its shadow. C) Back projection of original templates

The Retina Tracker continuously estimates a 4DOF transformation (rotation, scaling and translation) between current ROI and an internal planar map of the retina, the content of which is updated after each processed video image. The motion of the retina in the images is computed by tracking a structured rectangular grid of 30x30px templates equally spaced by 10px (see Fig. 4). Assuming that rotations and scale are small between image frames, the translation of individual templates visible (g) within the ROI is tracked by a local exhaustive search using Normalized Cross Correlation as the illumination invariant similarity metric (C_{gj}). In the following equations I is the input test image and T is the reference template image, both of the same size, while C_{gj} refers to the jth match in the local neighborhood of a visible template g. The metric operates on the three color channels (RGB) and is calculated as shown below:

$$C_{gj} = NCC_R + NCC_G + NCC_B \text{ where NCC}_v \text{ is :} \tag{1}$$

$$NCC_v = \frac{\sum_x \left(I_{x,v} - \overline{I_v} \right) \cdot \left(T_{x,v} - \overline{T_v} \right)}{\sqrt{\sum_x \left(I_{x,v} - \overline{I_v} \right)^2} \cdot \sqrt{\sum_x \left(T_{x,v} - \overline{T_v} \right)^2}}, \text{ where } x \text{ are pixels} \tag{2}$$

The confidence for each template (C_g) match is calculated by

$$C_g = \frac{\max(C_{gj})}{\overline{C_{gj}}} \tag{3}$$

For each template, we store its translation (P_g) and corresponding matching confidence which is used to improve robustness when matching in areas of minimal texture variation.

Template matching results are used as inputs in the iterative computation of the 2D rigid transformation from the image to the retinal map. In order to achieve real-time performance considering scaling, a Gaussian pyramid is implemented. The algorithm starts processing in the coarsest scale and propagates the results toward finer resolutions. At each iteration, the following steps are executed (see Fig. 5):

A) First, the average of motion for all visible templates g, weighted by their respective matching confidence (C_g), is used to determine the gross translation (P_i) relative to the reference templates' positions $\overset{\bullet}{P_g}$.

$$\vec{P_i} = \frac{\sum_g C_g (\vec{P}_g - \vec{P}_g^{\,\circ})}{\sum_g C_g} \tag{4}$$

B) Next, the gross rotation is computed by averaging the rotation (R_i) of the new template locations about the new origin (P_i), again weighted by the confidence:

$$R_i = \mathrm{atan2}\left(\sum_g C_g \cdot \sin\alpha, \sum_g C_g \cdot \cos\alpha\right), \text{ where :}$$

$$\alpha = \mathrm{atan2}\left(\vec{P}_{g,y} - \overline{\overline{P}}_{g,y}, \vec{P}_{g,x} - \overline{\overline{P}}_{g,x}\right) - \mathrm{atan2}\left(\vec{P}_{g,y}^{\,\circ} - \overline{\overline{P}}_{g,y}^{\,\circ}, \vec{P}_{g,x}^{\,\circ} - \overline{\overline{P}}_{g,x}^{\,\circ}\right) \tag{5}$$

C) finally, the scale (magnification) S_i is computed by comparing the average distance of template locations from the origin of the visible subset of the templates on the retina map and the current image:

$$S_i = \frac{\sum_g C_g \left\| \vec{P}_g - \overline{\overline{P}}_g \right\|}{\sum_g C_g \left\| \vec{P}_g^{\,\circ} - \overline{\overline{P}}_g^{\,\circ} \right\|} \tag{6}$$

At the end of each iteration, the target template grid positions are back-projected using the new transform (P_i, R_i, S_i ➔ T_i), and compared with those positions found in

the NCC matching step. The overall registration error (E_p) is calculated as the sum of squared position errors divided by the number of visible templates. The loop terminates when the registration error is below a predefined threshold e, which was chosen empirically to account for environmental conditions and retinal texture. In each iteration, the confidence (C_g) of the templates with high alignment errors (outliers) is reduced in proportion to the corresponding registration error for that template (E_g):

$$C_g = C_g - E_g \tag{7}$$

The algorithm favors contributions from templates that poses high template matching confidence and conform fairly closely to the regular grid. It was observed that this approach removes outliers in a few iterations. And, in the case where there are only a few templates with high confidence left, the registration result is considered unreliable.

We found this decoupled iterative method to be more reliable in practice than standard weighted least-squares approach. Outliers usually occur in areas where accurate image displacement cannot be easily established due to specularities, lack of texture, repetitive texture, slow color or shade gradients, occlusion caused by foreground objects, multiple translucent layers, etc. This also implies that any surgical instruments in the foreground are not considered in the frame-to-frame background motion estimation, making the proposed tracker compatible with intraocular interventions (see Fig. 4B). In the case of stereo images, rotation and scale of the left and right retina tracker as well as their vertical disparity are constrained to be the same (averaged) at each iteration of the algorithm.
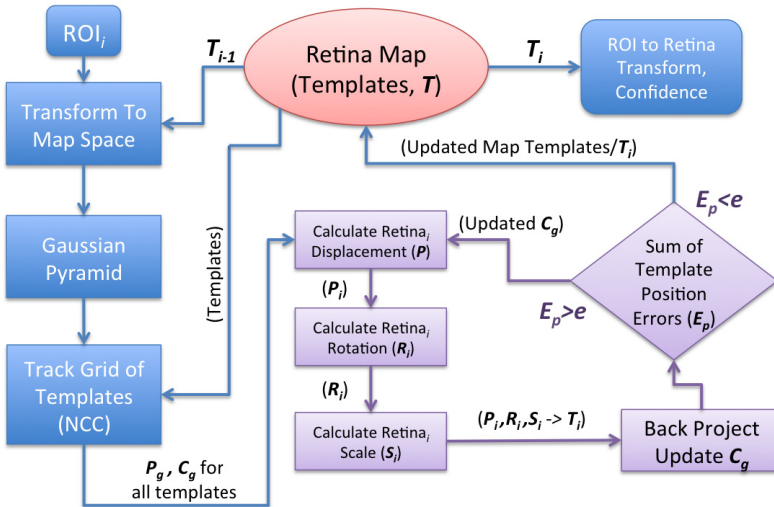

Fig. 5. Retina Tracker algorithm processing single video frame

The OCT Tracker provides the precise localization of the OCT beam projection on the retina which is essential for correlating OCT data with the anatomy. To facilitate robust tracking, we chose a camera sensor that captures OCT's near IR light predominantly on its blue RGB channel, as blue hues are uncommon in the retina (see Fig. 3A). The tracking algorithm (see Fig. 6) begins by thresholding the ROI in YUV color space to detect the blue patch. The area around this patch is then further segmented using adaptive histogram thresholding (AHT) on the blue RGB channel. Morphological operations are then used to remove noise from the binary image. This two-step process eliminates false detection of the bright light pipe, specular reflections on the metal shafts, and also accounts for common illumination variability. The location of the A-Scan is assumed to be at the centroid of this segmented blob. Initial detection is executed on the whole ROI while subsequent inter-frame tracking is performed within a small search window centered on the previous result. Left and right image tracker results are constrained to lie on the same image scan line.
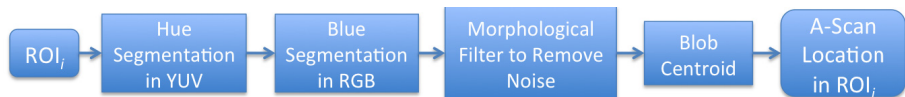

Fig. 6. OCT projection tracking algorithm

The Tool Tracker: In order to review past M-Scans with a standard surgical instrument, an existing visual tool tracking method for retinal surgery was implemented based on the work by Richa et al[13]. Like the OCT tracker, it operates on the ROI images and generates the tool pose with respect to the ROI of the retina. The algorithm is a direct visual tracking method based on a predefined appearance model of the tool and uses the sum of conditional variance as a robust similarity measure for coping with illumination variations in the scene. The tracker is initialized in a semi-manual manner by positioning the tool in the center of the ROI.

3   Experiments and Results

To evaluate the overall tracking performance we developed realistic water filled eyeball (25 mm ID) phantom. The sclera is cast out of soft silicone rubber (1mm thick near the lens), with an O-Ring opening to accept a surgical contact lens to simulate typical visual access, see Fig. 2. Two surgical trocars are used for tool access. The visual field of view is ~35 degrees considering a 5mm iris opening, which is comparable to that of a surgical case where ~20-45 degree vitreoretinal contact lenses are used. The eye rests in a plastic cup filled with methyl cellulose jelly to facilitate rotation. A thin, multi-layer latex insert, with hand painted vascular patterns, approximates the retina. These vascular details are coarser than those found in the human retina, but are still sufficient for tracking development, although not as good as the human retina's finer textures. Qualitative assessment by experienced vitreoretinal surgeons verified the ability of the model to simulate realistic eye behavior in surgical conditions. For independently verifiable ground truth ERM model, we chose a ~1mm

sliver of yellow, 60μm thick, polyester insulation tape which was adhered to the surface of the retina. It is clearly visible in the video images and its OCT image shows high reflectivity in comparison with the less intense latex background layers.
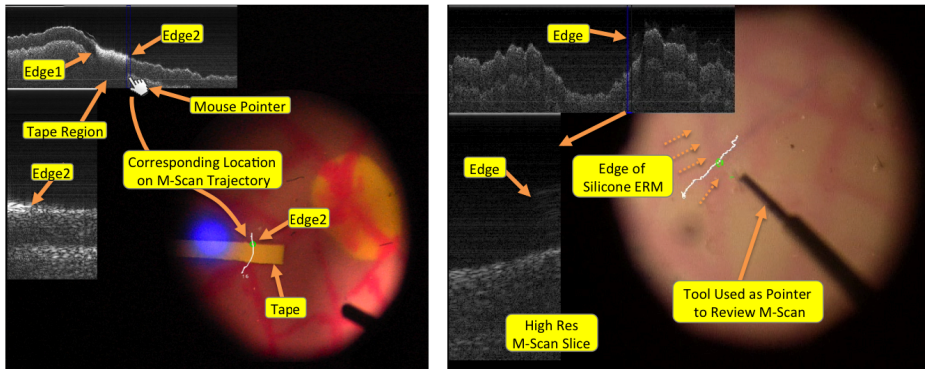

Fig. 7. A) M-Scan in Eye Phantom with tape simulating ERM used for validation of overall tracking.   B) M-Scan with silicone layer (invisible ERM) demonstrating more realistic surgical scenario. The surgeon uses the forceps as a pointer to review the M-Scan. The green circle is the projection of the pointer on the M-Scan path and corresponds to the location of the blue line on the OCT image and the zoomed-in high-resolution OCT "slice" image on the left.

To assess the overall accuracy of the system, 15 M-Scans were performed in the following manner: the area near the tape was first explored by translating the eye to build an internal map of the retina. Then, the OCT probe was inserted into the eye and an M-Scan was performed with a trajectory shown in Fig. 7A. The location of the edges of the tape was explored using a mouse pointer on the OCT image. The corresponding A-Scan location was automatically highlighted on the scan trajectory. The captured video of the display was then manually processed to extract the pixel location of the tape edge for comparison with the location inferred from the M-Scan. The average overall localization error, which includes retina and OCT tracking, was 5.16 ± 5.14px for the 30 edges analyzed. Considering an average zoom level in this experiment, this error is equivalent to ~100μm, using the tape width (~55px) as a reference. Largest errors were observed when the scan position was far from the retina map origin. This is mainly due to the planar approximation model map, as well as distortions caused by the lens periphery. With higher magnifications this error is expected to decrease.

To independently validate the OCT tracker, 100 image frames were randomly chosen from the experimental videos. The position of the OCT projection was manually segmented in each frame and compared to the OCT tracking algorithm results, producing an average error of 2.2 ± 1.74px. Sources of this error can be attributed to manual segmentation variability, as well as OCT projection occlusions by the tool tip when the tool was closer than ~500μm to the retina.

Additionally, for the purpose of demonstration a thin layer of pure silicone adhesive was placed on the surface of the retina to simulate a scenario where an ERM is

difficult to visualize directly. Fig. 7B shows the enface image of the invisible membrane and the corresponding M-Scan disclosing its cross-sectional structure. The surgeon can use the M-Scan functionality to determine the extents of the ERM and use the edge location to begin peeling.

4   Discussion

In this paper we presented a prototype for intraocular localization and assessment of retinal anatomy by combining visual tracking and OCT imaging. The surgeon may use this functionality to locate peeling targets, as well as monitor the peeling process for detecting complications and assessing completeness, potentially reducing the risk of permanent retinal damage associated with membrane peeling. The system can be easily extended to include other intraocular sensing instruments (e.g. force), can be used in the monitoring of procedures (e.g. laser ablation), and can incorporate preoperative imaging and planning. The methods are also applicable to other displays such as direct image injection into the microscope viewer presented in [14].

Our system can help a surgeon to identify targets, found in the OCT image, on the surface of the retina with the accuracy of   ~100 ± 100µm. This can easily be improved by increasing microscope magnification level or by using higher power contact lens. These accuracy values are within the functional range for a peeling application where the lateral size of target structures, such as ERM cavities, can be hundreds of microns wide, and the surgeons are approaching their physiological limits of precise freehand micro-manipulation [15]. We found that the retina tracking is the dominant component (~60%) of the overall tracking error due to high optical distortions and the use of the planar retina model. Since the retinal model does not account for retinal curvature, the background tracker is only reliable when the translations of the retina are smaller than ⅓ of the ROI size. Furthermore, preliminary in-vivo experiments on rabbits are very encouraging, showing similar tracker behavior as in the eye phantom. Additionally, the system does not include registration between tracking sessions, i.e. when the light is turned on and off. These considerations will be addressed in the future with feature-based registration methods and higher order retina model.

We are developing a new OCT probe to image at greater distances from the retina (e.g. 3mm), which will decrease the risk of accidental collisions and improve OCT tracking precision. We are currently reducing hand tremor-related OCT image instability through signal processing, as well as the robot assisted scanning methods presented in [6].

The promising results of this single phantom study demonstrated the system's novel function and its potential and are the basis for further extensive multi-phantom, multi-user study. This study will also provide qualitative data for the redesign of the user interface, which, for this prototype, was developed in cooperation with our surgeon co-authors.

References

1. Wilkins, J.R., Puliafito, C.A., Hee, M.R., Duker, J.S., Reichel, E., Coker, J.G., Schuman, J.S., Swanson, E.A., Fujimoto, J.G.: Characterization of epiretinal membranes using optical coherence tomography. Ophthalmology 103(12), 2142–2151 (1996)
2. Hirano, Y., Yasukawa, T., Ogura, Y.: Optical coherence tomography guided peeling of macular epiretinal membrane. Clinical Ophthalmology 5, 27–29 (2011)
3. Ehlers, J.P., Tao, Y.K., Farsiu, S., Maldonado, R., Izatt, J.A., Toth, C.A.: Integration of a Spectral Domain Optical Coherence Tomography System into a Surgical Microscope for Intraoperative Imaging. IOVS 52, 3153–3159 (2011)
4. Fleming, I.N., Voros, S., Vagvolgyi, B., Pezzementi, Z., Handa, J., Taylor, R., Hager, G.D.: Intraoperative Visualization of Anatomical Targets in Retinal Surgery. In: IEEE Workshop on Applications of Computer Vision, WAC 2008, pp. 1–6 (2008)
5. Han, S., Sarunic, M.V., Wu, J., Humayun, M., Yang, C.: Handheld forward-imaging needle endoscope for ophthalmic optical coherence tomography inspection. J. Biomed. Opt. 13(2), 020505 (2008)
6. Balicki, M., Han, J.-H., Iordachita, I., Gehlbach, P., Handa, J., Taylor, R., Kang, J.: Single Fiber Optical Coherence Tomography Microsurgical Instruments for Computer and Robot-Assisted Retinal Surgery. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 108–115. Springer, Heidelberg (2009)
7. Mountney, P., Giannarou, S., Elson, D., Yang, G.-Z.: Optical Biopsy Mapping for Minimally Invasive Cancer Screening. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 483–490. Springer, Heidelberg (2009)
8. Yamamoto, T., Vagvolgyi, B., Balaji, K., Whitcomb, L.L., Okamura, A.M.: Tissue property estimation and graphical display for teleoperated robot-assisted surgery. In: ICRA 2009, pp. 4239–4245 (2009)
9. Atasoy, S., Mateus, D., Meining, A., Yang, G.Z., Navab, N.: Endoscopic Video Manifolds for Targeted Optical Biopsy. IEEE Transactions on Medical Imaging (November 2011)
10. *cisst-saw* software library: https://trac.lcsr.jhu.edu/cisst
11. Liu, X., Balicki, M., Taylor, R.H., Kang, J.U.: Towards automatic calibration of Fourier-Domain for robot-assisted vitreoretinal surgery. Opt. Express 18, 24331–24343 (2010)
12. Liu, X., Kang, J.U.: Progress toward inexpensive endoscopic high-resolution common-path OCT. In: Proc. SPIE, vol. 7559, p. 755902 (2010)
13. Richa, R., Balicki, M., Meisner, E., Sznitman, R., Taylor, R., Hager, G.: Visual Tracking of Surgical Tools for Proximity Detection in Retinal Surgery. In: Taylor, R.H., Yang, G.-Z. (eds.) IPCAI 2011. LNCS, vol. 6689, pp. 55–66. Springer, Heidelberg (2011)
14. Berger, J.W., Madjarov, B.: Augmented Reality Fundus Biomicroscopy. Arch. Ophthalmol. 119 (December 2001)
15. Riviere, C.N., Jensen, P.S.: A study of instrument motion in retinal microsurgery. In: Proc. Int. Conf. IEEE Engineering in Medicine and Biology Society, pp. 59–60 (2000)

Real-Time Feature Matching for the Accurate Recovery of Augmented-Reality Display in Laparoscopic Videos

Gustavo A. Puerto-Souza, Alberto Castaño-Bardawil,
and Gian-Luca Mariottini

Department of Computer Science and Engineering,
University of Texas at Arlington,
416 Yates Street, 76019 Texas

Abstract. Augmented-Reality (AR) displays increase surgeon's visual awareness of high-risk surgical targets (e.g., the location of a tumor) by accurately overlaying pre-operative radiological 3-D model onto the intra-operative laparoscopic video. Existing AR systems are not robust to sudden camera motion or prolonged occlusions, which can cause the loss of those anchor points tracked along the video sequence, and thus the loss of the AR display. In this paper, we present a novel AR system, integrated with a novel feature-matching method, to automatically recover the lost augmentation by predicting the image locations of the AR anchor image-points after sudden image changes. Extensive experiments on challenging surgical video data are presented that show the accuracy, speed, and robustness of our designs.

1   Introduction

Intra-operative Augmented Reality (AR) systems are appealing in robotic-assisted laparascopic surgery, because of their capability to increase surgeons' awareness of high-risk surgical targets. By accurately overlaying a patient's pre-operative radio-logical (e.g., CT) 3-D data onto the live surgical video, AR systems can be helpful to reveal the position and orientation of an organ (c.f., Fig. 1), or of other important surgical targets (e.g., a tumor).

In order to ensure an accurate overlay, AR systems must be able to detect and reliably track in the laparoscopic video the *anchor points*, i.e., those matching features between the 3-D CT model and the observed surface of the organ.
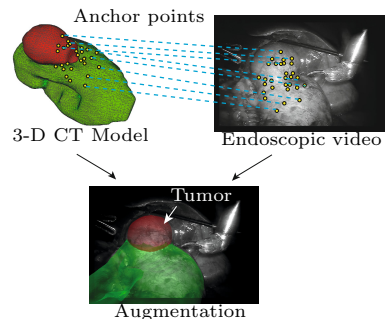

Fig. 1. *Example of an AR display*: A set of anchor-points (i.e., matches between the 3-D CT model and the endoscopic image) are used to maintain the AR display

Recent research efforts in the medical-imaging community have been devoted to improve the performance of feature tracking [1–5] in the cases of organ deformation and illumination changes. However, existing methods only work under the assumption that the features to track lie on *consecutive frames* and are then unable to recover those anchor points after *prolonged occlusions* (e.g., surgical tools passing in front of the camera), *rapid camera motions* (e.g., retractions), or *strong organ deformations.*

To the best of our knowledge, existing AR-displays pipelines can recover from the above occurrences only by means of a time-consuming manual user re-initialization. It is then of primary need to provide a solution that will *automatically* recover in *real time* the augmentation after the aforementioned events.

In this paper we propose to integrate our novel feature-matching HMA algorithm into an AR pipeline to automatically recover the precise position of a high number of anchor points. The HMA (Hierarchical Multi-Affine) feature-matching algorithm [6] predicts the position of the anchor points after a sudden image change, such as a prolonged and total camera occlusion. In this way, our method can be adopted to re-initialize the AR display, as well as any our feature tracker (c.f., Fig. 2).
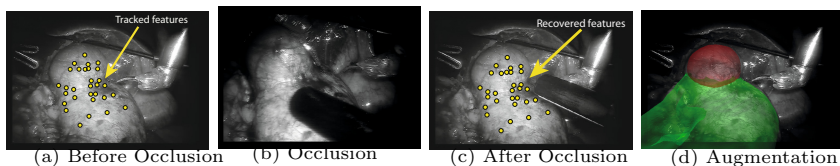

(a) Before Occlusion   (b) Occlusion   (c) After Occlusion   (d) Augmentation

Fig. 2. *Anchor-points recovery:* Feature-matching algorithms can be used to recover the location of anchor points in the image after a prolonged occlusion. (a): Features (dots) tracked during surgery; (b): Occlusion of ultrasound probe; (c) Recovered anchor points; (d) Augmented 3-D model.

Differently from feature tracking, our feature-matching method [6–9] does not make any assumptions about the sequential nature of the two input images (or about the observed scene) and is then preferred to address prolonged occlusions.

We validated our designs on a large set of laparoscopic video dataset from real partial-nephrectomy interventions. We first show that HMA can accurately retrieve the position of a large number of anchor points. In particular, we study the effectiveness of feature matching by comparing the efficiency of several image-features extractors (SIFT [7], SURF [10], ASIFT [11]), as well as three popular appearance-based matching techniques. We illustrate the effectiveness and computational time of each method, when used together with HMA, to recover anchor points. Our tests show that extracting SIFT features, combined with exhaustive appearance-based feature matching, allow HMA to recover a high percentage of anchor points, with low computational time, and high accuracy. Second, we present results from three challenging augmented-display videos, which show that our overall AR-recovery pipeline can achieve good augmented-reality accuracy, together with a high-number of recovered features.

2   Feature Matching for Tracking Recovery

The goal of *feature matching* consists of retrieving a set of similar features (*correspondences*) between two distinct images, e.g., before and after an occlusion. These features (e.g., SIFT [7], SURF [10], etc.) are usually extracted from both images and consist of a *keypoint* and of a *descriptor*, containing information about both the geometric properties (e.g., position) and the local image appearance around each keypoint [12], respectively.

We are here interested in using feature-matching strategies to leverage these sparse features to estimate mapping functions from one image to the other. These mappings will be indeed useful to predict the image location of the anchor points in the image after the occlusion. Feature-matching algorithms leverage first the appearance information contained in the feature descriptors to obtain a set of initial (candidate) matches [13,14] (c.f., Fig. 3(a)). Second, these initial matches are further pruned from possible initial ambiguities by means of *geometric constraints*, which leverage the geometric properties of each keypoint.

As mentioned above, this geometric constraint is particularly appealing since it can be used to *predict* the position of one (or a group of) anchor points in the image *after the occlusion* (c.f., Fig. 3(b)).
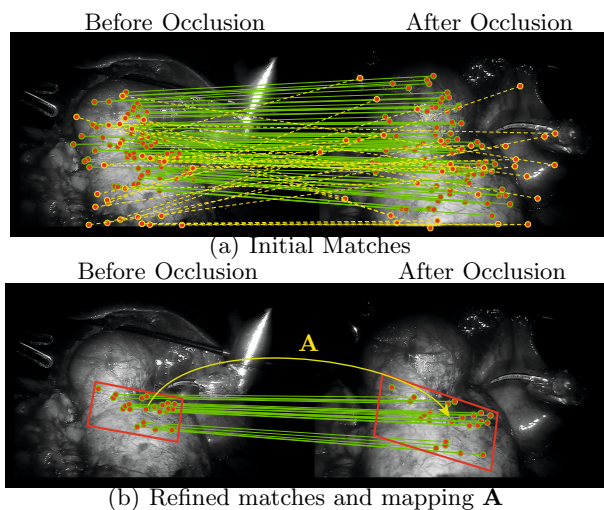

(a) Initial Matches

(b) Refined matches and mapping **A**

Fig. 3. *Feature matching for anchor-points retrieval:* (a) Local image appearance is used to find a set of initial matches (the correct matches are shown in (green) solid lines, and the incorrect matches in (yellow) dashed lines). (b) Subset of matches agreeing with a geometric constraint (affine mapping, **A**). Note that the use of a *single* affine transformation retrieves only a *few* matches localized *only* in a portion of the organ's surface (better seen in color).

An accurate estimation of these geometric constraints is then crucial for the success of the tracking recovery. The methods proposed in [6–9, 15] leverage the initial matches to estimate image mappings based on geometric constraints.

As a result, those potential matches that agree with this geometric model are considered as *inliers* (i.e., correct matches) and are not discarded as *outliers*.

In particular, the method in [15] generates virtual training images to estimate a single homography mapping together with the associated inliers. However, this method is computationally expensive and can only detect a few correspondences. The algorithm in [7] uses a unique affine-transformation constraint [16] and can reliably discard many wrong matches; however the detected inliers are limited mostly to a planar portion of the organ's surface. The above issue was solved in [9], where *multiple* local-affine mappings are used to detect a larger number of matches over the entire non-planar object surface. The work in [8] uses a dissimilarity measure and an agglomerative algorithm, to detect inliers and mutltple local transformations. However [9] and [8] are both slow for real time applications and do not scale properly for high numbers of initial matches.

In this work, we explore the integration of our recent Hierarchical Multi-Affine (HMA) feature-matching algorithm [6] to *accurately* detect and map a *high* percentage of accurate matches. Compared to [7,8], HMA is appealing because it can *accurately* retrieve the position of many anchor points over the entire organ's surface, and is *fast* because of the adoption of a hierarchical feature-clustering phase. This allows HMA to fulfill the speed, robustness, and accuracy requirements of a real-time AR surgical system.

2.1 Overview of HMA: Hierarchical Multi-Affine Feature Matching

HMA uses the set of initial sparse features (e.g., SIFT) to estimate *multiple* affine transformations, $\mathbf{A}_i$, spatially distributed over the entire organ's surface. As illustrated in Fig. 4, each affine mapping can be used to predict the location of anchor points in the image after the occlusion. HMA can simultaneously compute these transformations, as well as a set of feature matches (inliers) that support each affine transformation. Since in HMA these transformations locally adapt to the object surface, HMA can *i)* retrieve a larger number of correct matches when compared to existing strategies (e.g., [7]) and *ii)* can estimate a set of highly-precise image mappings. In what follows we present an overview of HMA. More details can be found in [6].
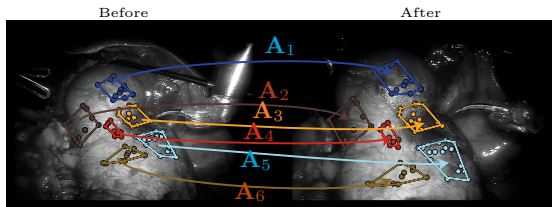

Fig. 4. From a set of initial (appearance-based) matches, HMA (1) retrieves a set of *refined (or final) matches*, and (2) estimates a set of *local affine transformations* (better seen in color)
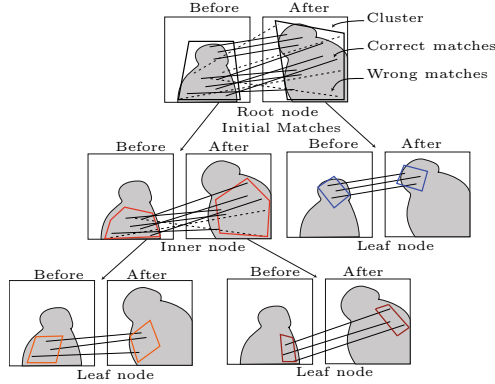
Fig. 5. *Illustrative example of HMA*: The initial matches (correct matches = solid lines; wrong matches = dotted lines) are hierarchically divided into smaller clusters (represented by polygons) each one limited to a portion of the image. Note that this phase generates clusters spatially distributed over all the organ's surface.

As illustrated in Fig. 5, the HMA algorithm uses a hierarchical k-means approach to cluster sparse image features. Each node of the tree structure represents a cluster of matches, and the edges represent the expansion of a cluster (node) into sub-clusters (children nodes).

Note that all the candidate (appearance-based) matches are contained in the root node (as indicated by the black polygon in the root node of Fig. 5). These matches are clustered into new nodes (see the colored polygons in the second level of the tree). For each node, geometric constraints are used to remove outliers. This clustering process continues until a criterion is reached (percentage of inliers in the node higher than a certain threshold). For example, the left ("inner node") in Fig. 5, is expanded because it contains many outliers, while the right node already reached a terminal state ("leaf node") since it does not have either any incorrect matches, nor a minimum number of supporting matches.

Each node (or group of features) is expanded by means of three steps: *(1)* a *clustering step*, which partitions the feature matches at that node into k clusters; *(2)* an *Affine-transformation estimation step* which robustly estimates an affine transformation, as well as the set of inliers. *(3)* Finally, a *correction step* is used to *(3.i)* verify that the clusters are spatially disjoint; to *(3.ii)* reassign the matches from one cluster to another to allow improved accuracy, and finally, to *(3.iii)* update the sets of inliers and outliers. After these stages, a *stop criterion* is adopted to evaluate when to stop the node expansion, and deem that a leaf node.

When all the branches are considered as leaf nodes, then the set of final transformations are estimated from the matches in each leaf node, and according to a threshold over a minimal number of inliers.

3   HMA for Augmented-Reality Recovery

We here present the integration of HMA into an augmented-reality pipeline. Initially, our AR system tracks image points in a laparoscopic video by means of a corner tracker [17]. These image points have been initially associated to 3-D points on the organ's model, and represent the anchor points. As an illustrative example of the AR-recovery scenario we will focus to the case of an organ occlusion caused by the motion of a surgical tool in front of the camera while the organ and the camera are also moving. In this case, the tracked features are lost and HMA is used to recover their position after such an event.

The block diagram in Fig. 6 illustrates the steps used to *automatically* and *accurately* recover the position of the anchor points in the image *after* the occlusion and, thus, to recover the augmented overlay.

Our method uses the image *before the occlusion* and the one *after the occlusion*. These two images are chosen automatically as those with the higher and lower number of tracked features, respectively, over the current video sequence. Note that, the position of the anchor points before the occlusion, i.e., the set of image features, $\{\mathbf{u}_i\}$ and the corresponding 3-D points, $\{\mathbf{X}_i\}$ from the CT organ's model, are assumed known initially (See Fig. 6).
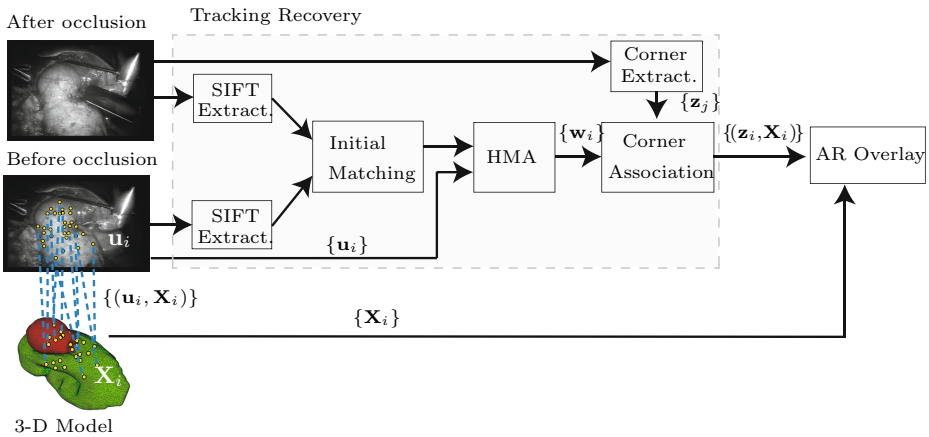

Fig. 6. *Diagram of the tracking recovery*: Our method uses the images *before* and *after* the occlusion, and initial associations (anchor points $\{(\mathbf{u}_i, \mathbf{X}_i)\}$) between the image before the occlusion and the 3-D model. As an output we recover the alignment between the image *after* the occlusion and the 3-D model.

First, SIFT features are extracted from both images (*SIFT Extraction* blocks) and are matched by means of an appearance-based criteria to find a set of candidate matches (*Initial Matching* block)[1]. These initial matches and the set of tracked points $\{\mathbf{u}_i\}$ (before occlusion) are used by the *HMA* block to predict the position of the tracked points, $\{\mathbf{w}_i\}$, in the image after the occlusion.

---

[1] A distance-ratio threshold of 0.9 is a common choice [7].

However, due to uncertainties and image noise, these predicted points might not precisely correspond to actual Harris corners [22] in the image after the occlusion. We have indeed found that re-initializing the feature tracker by using Harris corners guarantees a higher stability in the subsequent video frames. In order to refine the position of the predicted points, we first detect a set of Harris corners, $\{\mathbf{z}_j\}$ (*Corner Extraction* block) in the image after the occlusion. Second, we associate to each $\mathbf{w}_i$, the nearest neighbour corner $\mathbf{z}_j$. This is done in the *Corner Association* block. We use a threshold τ (e.g., 5 pixels) in order to preserve only the most certain associations. As a result, each point $\mathbf{z}_i$ will now *correspond* to each $\mathbf{u}_i$, so that a set $\{(\mathbf{z}_i, \mathbf{X}_i)\}$ of anchor-point associations is finally obtained. All these pairs are used to finally recover the augmented display in the image after the occlusion. Finally, the feature tracker is re-initialized using the corners $\{\mathbf{z}_i\}$, and the augmentation is kept until a new sudden image change occurs.

Note that some portions of the image before the occlusion might not be observable anymore in the image after the occlusion. In order to avoid the "recovery" of those tracked features that are not observed anymore after the occlusion, we limit the recovered points to the ones lying inside the convex-hull (of the feature positions) of the matches obtained from HMA.

4   Experimental Results

We experimentally validated our designs on a large set of endoscopic data obtained from real partial-nephrectomy interventions.

In particular, we first used a large and manually annotated dataset of image pairs to compare the performance of HMA in the case of different features and appearance-based matching methods. These quantitative tests allowed us to select the best combination of keypoint/descriptors, and matching technique. We believe this evaluation is of particular importance for the medical-imaging community because feature extraction and processing represents a crucial step for augmented reality. Secondly, we applied our augmented-reality prototype to three challenging surgical videos, and show that our feature-recovery pipeline achieves very good accuracy while retrieving a high percentage of features after unexpected events.

4.1   Comparison of Feature Extraction and Matching Methods

The performance of the HMA algorithm has been evaluated on a large and challenging laparoscopic-surgery dataset of 100 image pairs, with a resolution of 704×480, acquired from six real videos of partial-nephrectomy interventions. Our laparoscopic-image dataset contains many cases in which there is a real need to accurately retrieve a precise and large number of matches. These cases range from prolonged and complete camera occlusion (e.g., due to the surgical tool motion in front of the camera), camera retraction and reinsertion, sudden camera motion, and specular reflections. Figure 7 shows two sets of images that are

representative for some of these scenarios, and their corresponding phenomenon, such as occlusion, camera retraction and/or reinsertion, etc. (left: before occlusion; center: occlusion; right: after).
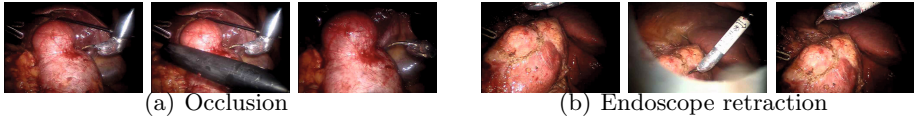

(a) Occlusion   (b) Endoscope retraction

Fig. 7. Representative images from the MIS dataset. (a) Three images showing a camera complete and prolonged occlusion and organ movement. (b) Camera retraction and re-insertion, with viewpoint change.

Our goal is to analyze which combination of feature keypoint/descriptor, as well as which appearance-based matching method, achieve HMA's best performance (consult [6] for a comparison between HMA and other state-of-the-art geometric feature-matching methods). In particular, we compared five keypoint detectors, two descriptor extractors and three matching techniques. The set of keypoint detectors consisted of three blob-based detectors (OpenCV's SIFT [7], SURF [18] and MSER [19]). As observed in Fig. 8(a), blobs feature detectors search for image regions that differ in intensity and texture properties [20]. In addition, we also used two corner-based detectors (OpenCV's FAST [21] and Harris [22]) that search the image for points with a large intensity variation in every direction [20] (c.f., Fig. 8(b)). Furthermore, three appearance-based feature-matching methods were evaluated: OpenCV's Brute-force Matcher (BF) with L1 (Manhattan Distance) and L2 (Euclidean Distance) norms, and the Fast Library for Approximate Nearest Neighbors (FLANN)-based matcher [23]. Note that BF performs an exhaustive search over the training and query image sets of keypoints to find possible match candidates, while FLANN performs an approximate search (using L2 norm) by means of search trees.
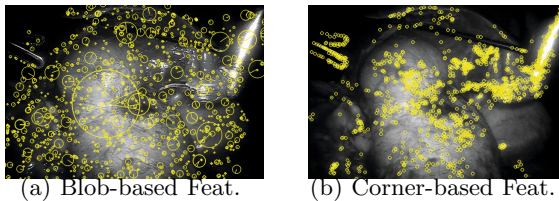

(a) Blob-based Feat.   (b) Corner-based Feat.

Fig. 8. (a) Example of a Blob-base Feature Detector. The yellow circles represent SIFT keypoints detected in an endoscopic image: the position of each feature is represented by the center of the circle, the scale is proportional to the radius, while the orientations contain the directions of the dominant local image gradients around the feature position. (b) Example of a Harris corner-based feature detector for the same image. Note that these features are only represented by their image position.

In order to provide an extensive evaluation over all of the above scenarios, manually-annotated ground-truth data were extracted from the aforementioned

image sets. This ground-truth data consists of a set of manually-selected corresponding corners for each image pair. Our dataset has an average of 19.07 ± 2.89 manually-selected corresponding corners. Note that selecting accurate ground-truth corresponding corners was indeed a very difficult task, especially because of the strong illumination changes, blur and clutter present in the images.

Our comparison is based on measuring the algorithm's accuracy in mapping ground-truth corresponding points between the two images (mapping performance). Additionally, we evaluated the computational time of the main three phases: the feature extraction, the appearance-based matching of descriptors, and HMA.

The computational time required for each task was measured in seconds of timer CPU (Intel^{TM}Core2 Duo 2.26GHz, Intel Corp., Santa Clara, CA). The experimental evaluation presented below analyzes and discusses the performance of HMA with all the aforementioned combinations of keypoints, descriptors and appearance-based matching strategies.

We initially compared *all the possible combinations* of keypoints/descriptors, with a fixed affine transformation precision (τ_c) of 5 pixels [6], in order to isolate those combinations that presented an unacceptable performance. These tests showed that the combinations (keypoint/descriptor) SIFT/SURF and HARRIS/SURF led to a poor outcome. This happens because these combinations produce appearance-based matches (initial matches) with very large percentage of wrong matches and only few (or none) correct matches. As a result, HMA failed (non-solution image pairs) for high percentage of the image pairs.

Similarly, other combinations of keypoint/descriptors such as SURF/SURF, SURF/SIFT, MSER/SURF, MSER/SIFT, and FAST/SURF presented high reprojection error values and high percentage rate of non-solution when evaluated under different appearance-based matching techniques. As a result, all the aforementioned combinations were discarded for further evaluation, and are not included because of the limited space and to focus our discussion on the most successful algorithms.

Similarly, FLANN-based feature matching presented the best performance in terms of computational time, however it did not show the same good results in terms of accuracy as for the exhaustive (Brute-force) methods. At the end, BF with norm L2 (BF-L2) was chosen because it obtained the best (lower) reprojection error values.

After these initial tests, we only preserved the following combinations of keypoint/descriptors: SIFT/SIFT, FAST/SIFT and HARRIS/SIFT. In the following we will name them only according to their feature detector since all of them use SIFT descriptors and BF matching. Their extensive performance comparison is reported in Table 1 (Feature Extraction and Appearance-based Matching), as well as in Table 2 (HMA accuracy).

From Table 1, we observed three interesting phenomena. First, the large variability on the number of detected features by FAST. Second, the column "time per feature" indicates that there is no significant statistical difference between the time required for each feature extractor. Third, note that even if FAST

Table 1. MIS Dataset Test: Feature Extraction and Appearance-based Matching

Local Features	Feature Extraction								App.Based Match.			
	No. Train.		No. Query		Time [s]		Time/feat. [s]		No.		Time [s]	
KP's/D's	avg	std	avg	std	avg	std	avg	std	avg	std	avg	std
SIFT/SIFT	395.4	103.8	1043.0	134.9	0.456	0.025	0.00032	0.00002	128.5	73.2	0.038	0.01
FAST/SIFT	587.0	336.1	1417.0	600.4	0.537	0.182	0.00028	0.00005	138.9	141.8	0.088	0.08
HARRIS/SIFT	343.1	199.2	689.0	287.9	0.238	0.020	0.00028	0.00013	90.3	78.4	0.025	0.02

*The extraction of features in the training image was limited to a region of interest

computed a larger number of features for each image, the number of appearance-based matches is very similar to SIFT.

Table 2 shows the results on the mapping accuracy for the three combinations and with different affine estimation precisions (5, 10 and 15 pixels). We observed that in general SIFT exhibits good results by consistently retrieving a high number of features with a lower reprojection error and low failure rate, while FAST and Harris obtain larger reprojection errors and larger failure rate. We consider SIFT obtained better results because of its invariance to scale and rotation. Conversely, FAST and Harris features not only lack of orientation and scale, they also rely on high textured areas.

Observe that FAST and HARRIS present considerably high percentages of retrieved inliers over the set of appearance-based matches (e.g., FAST approx. 92%); however, also note that their mapping errors are larger than SIFT. This may happen because FAST and Harris features are mostly localized over the same image regions, while SIFT matches are distributed more uniformly over the image, even near-by textureless areas (recall Fig. 8).

In summary, The reprojection error and the quantity of image pairs without a solution for Harris are larger than the other methods. FAST presents higher percentage values of retrieved inliers but with larger reprojection errors and larger number of data without a solution. Therefore we can conclude that SIFT features with BF-L2 is the best choice, exhibiting an affine-transformation mapping precision in the range $5 - 10$ pixels. It also has the best performance by offering a high percentage of retrieved features (49.5%), low extraction time per feature $(0.32 \pm 0.02$ ms) and reduced number of non-solution image-pairs (14%).

4.2   Tracking Recovery and Augmented Display

We validated the performance of our tracking-recovery on three video sequences extracted from a partial nephrectomy intervention. In these sequences the anchor points are easily lost because of several occurrences, such as prolonged and complete camera occlusions due to *a)* the motion of an ultrasound probe in front of the endoscope, *b)* breathing motion, and *c)* sudden and fast camera and organ motion.

In all of these three sequences, we used a 3-D model of the kidney obtained by processing pre-operative CT-data from the same patient. In particular, ITK-Snap was used to segment the kidney and the tumor and to generate the 3-D models (MeshLab was used to smooth the surface). Without loss of generality,

Table 2. MIS Dataset Test: HMA Mapping Accuracy

| Local Features | τ_c | HMA Algorithm | | | | Reproj. Error | | | Non- |
| | | Inl. HMA | | Time [s] | | [pixels] | | | Sol |
KP's/D's		avg	std	avg	std	avg	std	med.	Img's
SIFT/SIFT	5	44.7%	23.5%	0.0215	0.0156	4.84	4.54	3.47	13%
	10	49.5%	22.9%	0.0211	0.0152	4.74	4.62	3.41	14%
	15	49.6%	22.9%	0.0207	0.0150	5.24	5.29	3.42	12%
FAST/SIFT	5	65.8%	16.9%	0.0588	0.0883	5.90	5.37	4.19	16%
	10	71.8%	16.9%	0.0578	0.0869	5.74	4.99	3.95	17%
	15	72.4%	16.9%	0.0570	0.0870	5.66	5.20	3.91	16%
HARRIS/SIFT	5	53.3%	23.6%	0.0258	0.0281	6.65	6.05	4.76	25%
	10	55.3%	24.1%	0.0242	0.0261	11.41	9.67	4.91	23%
	15	56.7%	22.2%	0.0240	0.0265	15.12	15.80	5.27	22%

*τ_c: affine transformation accuracy, in pixels.
*Inl. HMA: Percentage of HMA Inliers over total Matches (Initial Matches)
*Non-Sol Img's: Percentage of images without a solution

we assume that an expert user has provided the initial anchor points for each sequence, i.e., the association between some corners in the first frame of the video and the corresponding vertices of the 3-D model. These corners were then tracked throughout the video sequence by using the method in [17].

Our tracking-recovery algorithms started when either one of the two following conditions was verified: *(i)* the number of tracked points in the current frame dropped below 30; or *(ii)* when the number of tracked points was below 60 *and* the average reprojection error is larger than 6 pixels[2].

a) Ultrasound-probe occlusion: This sequence consists of 170 frames, each one with resolution 704×480. The kidney is *mostly* occluded by the ultrasound probe during frames 20-130 and 160-170. Figure 9(1st row) shows the sequence of laparoscopic images with the augmented display in the case that *no recovery strategy is adopted*. In the first frame, and due to the anchor points selected by the user, the augmented display nicely overlays with the image of the kidney. However, around frame 25 and due to the presence of the ultrasound probe in the bottom part of the kidney, the number of tracked features decrease sensibly (see Fig. 9(2nd row)) thus causing the augmentation to drift. This effect becomes more predominant around frame 100 for which the tracked features are very few and the augmentation is wrong, and eventually was lost right after.

The performance of our *tracking-recovery algorithm* are illustrated in Fig. 9(3rd row) which shows the sequence of 3-D augmented laparoscopic images exactly at the same time instants of the case without recovery. At frame 25, despite the presence of the ultrasound probe on the bottom part of the kidney, our strategy

---

[2] The reprojection error is defined as the norm of the difference between the tracked points and the 3D-model projected points, which are estimated using the current frame's alignment with the 3-D model.
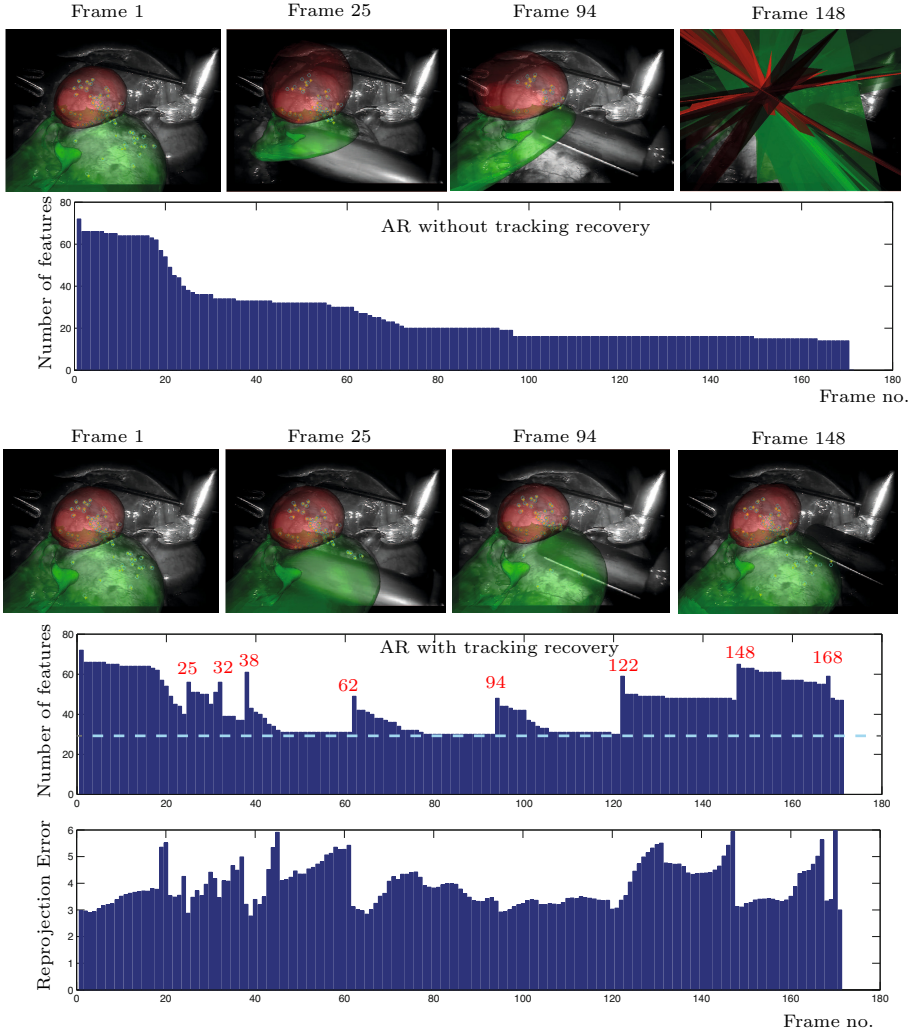
Fig. 9. *Tracking Recovery Results*: (1st row) AR system without the tracking recovery. (2nd row) The number of features tracked decreases without the recover phase; (3rd row) AR system with the tracking recovery is not stable; (4th row) Number of features per frame; we have identified those time instants in which a recovery phase took place. (5th row) Reprojection error never goes over 6 pixels.

successfully recovers several anchor points on the bottom of the kidney, thus resulting in a correct augmentation. Also, as illustrated in Fig. 9(4th row), the number of tracked features never goes under a user-selected minimum (30); note also that the peaks in the number of features correspond to those time instants for which the recovery was taking place and the augmentation was then successfully re-initialized. The final plot in Fig. 9(5th row) shows that the average reprojection error never

goes above a user-selected threshold (6 pixels in our case). The average reprojection error of the recovery was 4.01 ± 0.88 (pixels), and with a percentage of tracked corners of $62\% \pm 18\%$ with respect to the first frame.

b) Breathing Sequence: This sequence consists of 400 frames, in which the kidney was partially occluded between frames 250-380. In addition, this sequence is challenging due to fast organ motion caused by the patient's breathing. For this sequence, the average reprojection error of the recovery was 3.04 ± 0.43 (pixels), and with a percentage of tracked corners of $64\% \pm 12\%$ with respect to the first frame.

c) Fast Camera Motion Sequence: This sequence consists of 101 frames with a partial occlusion occurring at frames 75-101. In particular, this sequence presents a fast camera motion (towards the organ) which worsens the feature motion-tracking problem. The average reprojection error of the recovery was 3.31 ± 0.43, and with a percentage of tracked corners of $63\% \pm 12\%$.

5   Conclusions

This work presented the use of the HMA feature-matching algorithm to successfully recover the position of AR anchor points, and thus of the augmented display, that were lost due to unexpected image changes (e.g., occlusion). Our work represents an improvement over the current state-of-the-art AR systems, since it uses the HMA algorithm to successfully retrieve a high number of accurate anchor points. In addition, this is the first work to adopt feature-matching to successfully retrieve the lost anchor points and re-initialize the AR display in an effective way. The effectiveness of our tracking-recovery strategy has been demonstrated with an extensive laparoscopic dataset from real partial-nephrectomy intervention. We studied the effectiveness of feature matching by comparing the efficiency of several image-features extractors, as well as of three popular appearance-based matching techniques. We illustrate their accuracy and the computational time of each method when used together with HMA to recover anchor points. We observed that SIFT image-features together with Feature Matching method BF-L2 gave HMA the best performance, in terms of accuracy and efficiency. We also present results from three challenging augmented-display videos, which show that our overall AR-recovery pipeline achieves good augmented-reality accuracy, together with a high-number of recovered features.

References

1. Richa, R., Bo, A.P.L., Poignet, P.: Towards robust 3d visual tracking for motion compensation in beating heart surgery. Medical Image Analysis (2010)
2. Stoyanov, D., Mylonas, G.P., Deligianni, F., Darzi, A., Yang, G.Z.: Soft-tissue motion tracking and structure estimation for robotic assisted MIS procedures. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 139–146. Springer, Heidelberg (2005)
3. Mountney, P., Yang, G.-Z.: Motion compensated Slam for image guided surgery. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part II. LNCS, vol. 6362, pp. 496–504. Springer, Heidelberg (2010)

4. Giannarou, S., Visentini-Scarzanella, M., Yang, G.: Probabilistic tracking of affine-invariant anisotropic regions. IEEE T.PAMI (2012)
5. Yip, M.C., Lowe, D.G., Salcudean, S.E., Rohling, R.N., Nguan, C.Y.: Real-time methods for long-term tissue feature tracking in endoscopic scenes. In: Abolmaesumi, P., Joskowicz, L., Navab, N., Jannin, P. (eds.) IPCAI 2012. LNCS, vol. 7330, pp. 33–43. Springer, Heidelberg (2012)
6. Puerto-Souza, G.A., Mariottini, G.L.: Hierarchical multi-affine (HMA) algorithm for fast and accurate feature matching in minimally-invasive surgical images. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS (2012)
7. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comp. Vis. 60(2), 91–110 (2004)
8. Cho, M., Lee, J., Lee, K.M.: Feature correspondence and deformable object matching via agglomerative correspondence clustering. In: Proc. 9th Int. Conf. Comp. Vis., pp. 1280–1287 (2009)
9. Puerto-Souza, G.A., Adibi, M., Cadeddu, J.A., Mariottini, G.L.: Adaptive multi-affine (AMA) feature-matching algorithm and its application to minimally-invasive surgery images. In: 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2371–2376 (September 2011)
10. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
11. Morel, J.M., Yu, G.: ASIFT: A new framework for fully affine invariant image comparison. Journal on Imaging Sciences 2(2), 438–469 (2009)
12. Szeliski, R.: Computer vision: algorithms and applications. Springer-Verlag New York Inc. (2010)
13. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. 60, 1615–1630 (2005)
14. Mountney, P., Lo, B., Thiemjarus, S., Stoyanov, D., Yang, G.-Z.: A probabilistic framework for tracking deformable soft tissue in minimally invasive surgery. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 34–41. Springer, Heidelberg (2007)
15. Del Bimbo, A., Franco, F., Pernici, F.: Local shape estimation from a single keypoint. In: Proc. Comp. Vis. Patt. Rec. Workshops, pp. 23–28 (2010)
16. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge Univ. Press (2000)
17. Hailin, J., Favaro, P., Soatto, S.: Real-time feature tracking and outlier rejection with changes in illumination. In: Proc. IEEE Intern. Conf. on Computer Vision, vol. 1, pp. 684–689 (2001)
18. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
19. Urban, M., Matas, J., Chum, O., Pajdla, T.: Robust wide baseline stereo for maximally stable extremal regions. Image and Vision Computing 22(10), 761–767 (2004)
20. Siegwart, R., Nourbakhsh, I.R., Scaramuzza, D.: Introduction to Autonomous Mobile Robots. MIT Press (2011)
21. Rosten, E., Drummond, R.: Fusing points and lines for high performance tracking. In: Proceedings of the International Conference on Computer Vision, pp. 1508–1511 (2005)
22. Harris, C., Stephens, M.: A combined corner and edge detector. In: Alvey Vision Conference, Manchester, UK, vol. 15, p. 50 (1988)
23. Muja, M., Lowe, D.G.: Fast approximate nearest neighbors with automatic algorithm configuration. In: International Conference on Computer Vision Theory and Application, VISSAPP 2009, pp. 331–340. INSTICC Press (2009)

Automated Detection of Mediastinal Lymph Nodes for Assistance of Transbronchial Needle Aspiration

Takayuki Kitasaka[1], Mitsuhiro Kishimoto[2], Masahiro Oda[2], Shingo Iwano[3], and Kensaku Mori[4]

[1] School of Information Science, Aichi Institute of Technology
[2] Graduate School of Information Science, Nagoya University
[3] Graduate School of Medicine, Nagoya University
[4] Strategy Office, Information and Communications Headquarters, Nagoya University

Abstract. This paper proposes an automated method for detecting mediastinal lymph nodes from contrasted 3D chest CT images to assist transbronchial needle aspiration (TBNA). In lung cancer treatment, it is vital to check whether metastasis occurs in lymph nodes. For this purpose, two-phase contrasted CT images are used for precise and accurate image diagnosis in the clinical field. In the proposed method, two input CT images are first registered and then the processing area is calculated on the basis of intensity information and an atlas for the lymph node map. Lymph node metastasis candidates are extracted roughly by using local intensity structure analysis. Lymph node metastasis regions are obtained by using a shape modification process based on using the shape features of candidates to achieve region growth and false positive (FP) reduction. To provide assistance for TBNA, detection results are visualized through the semi-translucent bronchial wall by virtual bronchoscopy. In experimentally applying the method to 42 pairs of contrasted chest CT images, the results show that the method detected 92.9% of the lymph node metastases, while a radiologist detected 57.1%. The lymph nodes were clearly visualized in the virtual bronchoscopy as the simulation of TBNA guidance.

1 Introduction

Lung cancer is the most common cause of cancer-related death worldwide today. Early diagnosis and proper treatment are important to improve the survival rate. If lung cancer is diagnosed, it is necessary to resect not only the cancer but also enlarged lymph nodes, which may be metastases of the cancer. This is because lymph node metastases can be formed at parts remote from the cancer. Thus, image diagnosis of metastases plays a quite important role in lung cancer treatment [1]. To make diagnosis more precise, contrasted CT images are used which are taken especially in two phases, early- and late-phases. The development of computer assisted detection of enlarged lymph nodes is expected.

In addition, visualization of mediastinal lymph nodes is helpful for transbronchial needle aspiration (TBNA) guidance.

There are several state-of-the-art lymph node detection methods, such as the lymph node atlas-based and the lymph node model-based methods [2][3]. Feuerstein et al. proposed an atlas-based method of detecting lymph nodes from 3D chest CT images, including contrasted and non-contrasted images. The method utilizes blob structure enhancement filters to enhance rounded shadows and detects lymph node regions on the basis of a lymph node map [2]. Feulner et al. presented a lymph node detection method based on a classifier that has learned true and false positive samples detected by a segmentation algorithm. Final detection results are obtained by utilizing the graph cuts method. However, the method [2] detects many false positives (FPs), which leads to poor detection specificity. The occurrence of FP regions is induced by setting inappropriate processing areas, which are computed from a model (lymph node atlas) and image (edge information). Though the method [3] reduced the FPs number significantly, the true positive (TP) rate is still low for clinical application. In lymph node detection, it is important to improve the TP rate while keeping the FP rate low. Furthermore, these methods did not utilize the two-phase contrasted CT images commonly used in the clinical field.

In this paper we propose an automated method for detecting mediastinal lymph nodes by utilizing two-phase contrasted CT images, which are used for diagnosing metastases as mentioned above. We also visualize the detected lymph nodes by virtual bronchoscopy for TBNA guidance. We evaluate the proposed method by using 42 cases of CT image pairs and compare the results obtained with it with diagnosis results obtained by a radiologist. In Section 2, details of the lymph node metastasis detection method are provided. Section 3 shows experimental results obtained with 42 pairs of CT images (84 CT volumes). These are also compared with detection results obtained by a radiologist. We add a discussion about lymph node detection from both technical and clinical points of view.

2   Method

2.1   Overview

We detect and visualize lymph node metastasis regions from two-phase contrasted CT images by following four steps: (1) preprocessing, (2) rough extraction of lymph node candidates, (3) precise detection of lymph nodes, and (4) visualization of lymph nodes. In the preprocessing step, we normalize the intensity of the blood vessels and determine the processing area on the basis of a lymph node atlas. The rough extraction process detects lymph node region candidates by filtering. After these processes, we precisely detect lymph node regions by performing shape modification and FP reduction processes. Here, we denote the early- and late-phase CT images as I_E and I_L, respectively. The detailed procedures are given below.

(a) Illustration of the lymph node map.   (b) 3D display of lymph nodes.

(c) An example of axial slices. The red line indicates the contour of enlarged lymph nodes.
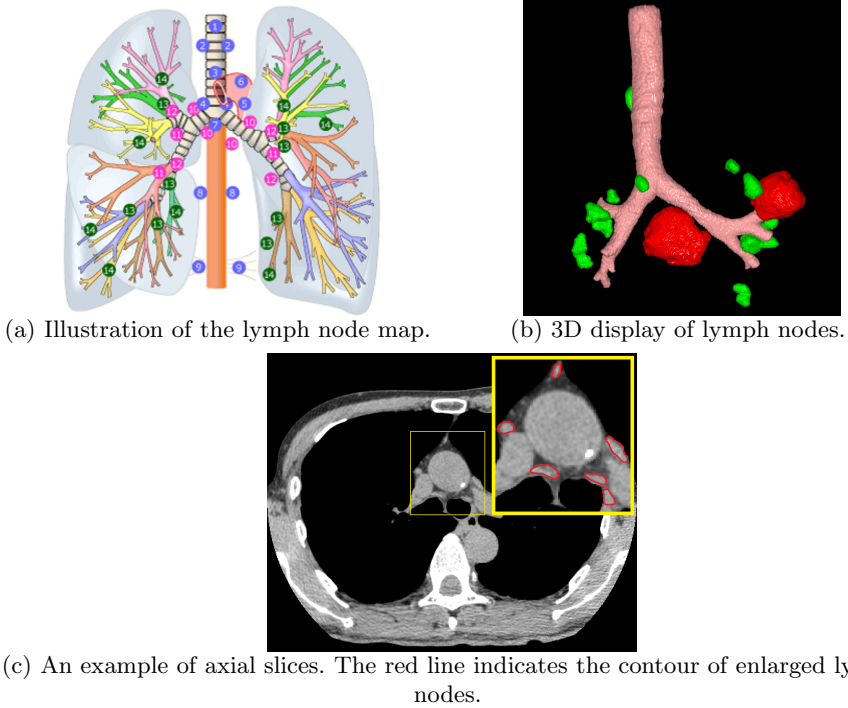
Fig. 1. Lymph node map and 3D display of lymph nodes from an actual medical image. (a) Illustration of the lymph node map and (b) an example of 3D display of lymph nodes (Pink: the airway, Red: lymph node metastases, and Green: normal enlarged lymph nodes), and (c) an example of axial slices of a contrasted chest CT image.

2.2 Preprocessing

Both I_E and I_L are smoothed by a median filter to remove image noise. I_E is then registered to I_L by using a non-rigid registration algorithm [4]. We denote the registered image as I_{El}. The aorta region is extracted from I_{El} by using Feuerstein's approach [5]. The mean intensity of blood vessels μ and its standard deviation σ are estimated by obtaining those values in the aorta region. Then, we extract a high intensity region R_H whose intensity is greater than $\mu - 3\sigma$. To remove the high intensity areas in I_L, which may cause FPs in the lymph node candidate extraction process described below, the intensity of R_H in I_L is replaced with -1000 H.U. (air intensity), denoting the image $I_{L'}$. The lymph node map atlas A, which is generated from the training dataset by using the previous method [2], is also registered to I_L. The atlas A contains the information for the number of lymph node stations.

R_T (Step a) R_A (Step b) R_P (Step c)
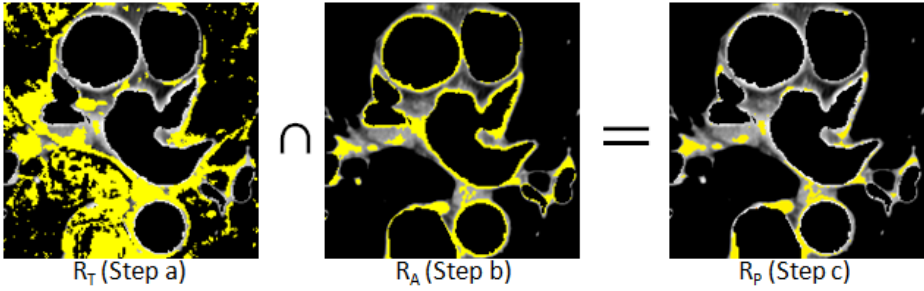
Fig. 2. Examples of the processing area rectriction

2.3   Rough Extraction of Lymph Node Candidates

We first restrict the processing area for the lymph node candidate extraction (Fig. 2). Since lymph node regions take higher intensity in $I_{L'}$ than in I_{El} due to the effect of the contrast medium, (Step a) a temporal region R_T is extracted from the image $I' = I_{L'} - I_{El}$ by thresholding against I' with the threshold 0 H.U. (water intensity). Also from the atlas information, (Step b) we extract the region R_A whose intensity is greater than the mean intensity of A in $I_{L'}$. Then, (Step c) the processing area R_P is obtained as $R_T \cap R_A$, followed by the closing operation whose structure element is a sphere with 1.0 mm radius.

We enhance blob-like intensity structures by analyzing the Hessian matrix of each voxel in R_P [6]. To detect lymph nodes of various sizes, the enhancement is performed in multi-scale of $\sigma_h = 1.5$ and 2.5. We extract voxels whose smallest eigenvalue λ_3 of the Hessian matrix is less than zero. This is a weak constraint in order to detect rounded objects. After the opening operation whose structure element is a sphere with 1.0 mm radius, the lymph node candidate region R_C is obtained (Fig. 3 (a)).

2.4   Precise Detection of Lymph Nodes

Shape Modification by Region Growing. First, an edge-preserving smoothing filter is applied to the image $I_{L'}$ [7]. Then, we apply a region growing algorithm to each connected component of R_C to modify each candidate shape. This is because the candidates detected by the rough extraction process cover only the central parts of lymph nodes. We try to extract precise lymph node regions by using a region growing process. The structure element of the region growing is set to a sphere with 1.5 mm radius. The growing conditions for a voxel of interest (i, j, k) are as follows.

[**Region growing conditions**]
    (a) $I_{L'}(i, j, k) > 0$,
    (b) $|Lap * I_{L'}(i, j, k)| < T_L$, where Lap is the Laplacian operator and T_L
        a threshold value,
    (c) Euclidean distance between (i, j, k) and the center point of each connected component is less than 15 mm.
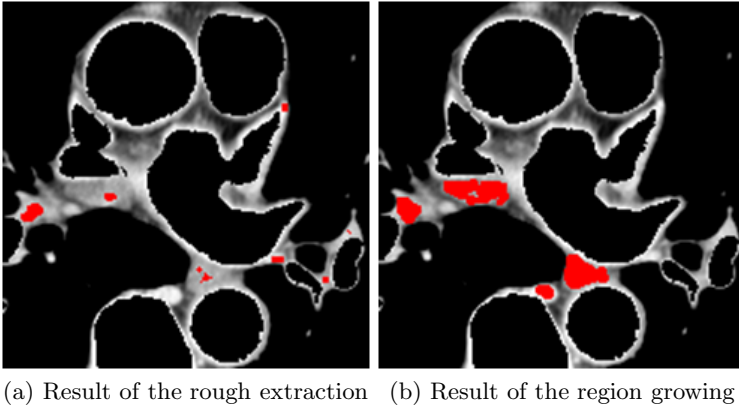
(a) Result of the rough extraction   (b) Result of the region growing

Fig. 3. Examples of rough and precise detection results

After the growing process, we obtain the modified candidate region R_G (Fig. 3 (b)).

FP Reduction Based on Shape. We eliminate FP regions by using the following procedure. First, the opening operation, whose element shape is merely a voxel, is applied to R_G to separate connected candidates after the region growing. In the clinical field, it is important to detect lymph nodes with diameter of 10 mm or larger. Hence, we eliminate small candidates that satisfy any of the following conditions:

[**FP reduction condition**]
  (i) the length L_{min} along the smallest axis of the candidate is less than T_S mm,
  (ii) the length L_{max} along the longest axis of the candidate is greater than 60.0 mm,
  (iii) $L_{max}/L_{min} > T_F$,
  (iv) $L_{max} > 30$ [mm] and $L_{max}/L_{min} > 2$.

Here, the smallest and longest axes are calculated by using the PCA technique. T_S and T_F are threshold values determined empirically. Conditions (ii), (iii), and (iv) eliminate the blood vessel and esophagus regions. Then, candidates that do not overlap with R_A at all are also eliminated. The remaining regions are detected as the lymph node metastasis regions.

2.5   Visualization of Lymph Nodes

In TBNA visualization, it is important to clarify where lymph nodes exist from the bronchial lumen, namely with a virtual bronchoscopic camera. The bronchial wall is rendered as a semi-translucent area through which mediastinal lymph nodes can be seen. In addition, lymph nodes are colored on the basis of the number of lymph node stations by using the atlas information.

3 Experiments and Results

3.1 Experiments

We applied the proposed method to 42 pairs of contrasted 3D chest CT images. The acquisition parameters of the images are: 512×512 pixels, 338-469 slices, 0.625-0.782 mm pixel spacing, and 0.799-0.801 mm slice spacing. The parameters used in the experiment are $T_L = 500$, $T_S = 4.0$, and $T_F = 3.0$. We conducted two kinds of evaluations: (**Accuracy evaluation**) for lymph nodes whose diameter is greater than 10.0 mm, regardless of whether they are metastases or not, and (**Clinical evaluation**) for metastasis lymph nodes that were pathologically identified. The sensitivity and the specificity used for evaluation criteria were calculated by TP/(TP+FN) and TP/(TP+FP), respectively. We compared the results obtained with the proposed method, those obtained with the previous method [2], and a radiologist's detection results.

3.2 Results

Table 1 shows the **Accuracy evaluation** results obtained with the previous and proposed methods. The proposed method was able to detect 89.2% of the enlarged lymph nodes, while the previous method detected 91.4%. The sensitivity of the proposed method was slightly lower but the specificity was about 13 % higher (60/case to 10/case). Table 2 shows the **Clinical evaluation** results. The proposed and previous methods were able to detect 92.9% of the lymph node metastases, while the radiologist detected 57.1%. The FP number obtained with the proposed method was about 50/case lower than that obtained with the previous method. Though the sensitivity obtained by the radiologist was lower than that obtained by automated detection, the specificity was quite high, i.e., most of the lymph nodes pointed out by the radiologist were truly metastases.

Figure 4 shows examples of the detection results obtained by using the previous and proposed methods. As the figure shows, the proposed method had a smaller number of FPs. The lymph node shapes detected by using the proposed method were more correct than those obtained by using the previous one. Figure 5 shows examples of lymph node metastases that both the proposed method and the radiologist detected, and examples of those that the proposed method detected but the radiologist missed. Figure 6 shows examples of those that both the proposed method and the radiologist missed. Fifteen of the sixteen metastases detected by the radiologist were also detected by using the proposed method. Eleven of the twelve missed by the radiologist were detected by using the proposed method. Figure 8 indicates examples of the mediastinal lymph nodes visualized by virtual bronchoscopy. The nodes were well visualized through the semi-translucent bronchial wall.
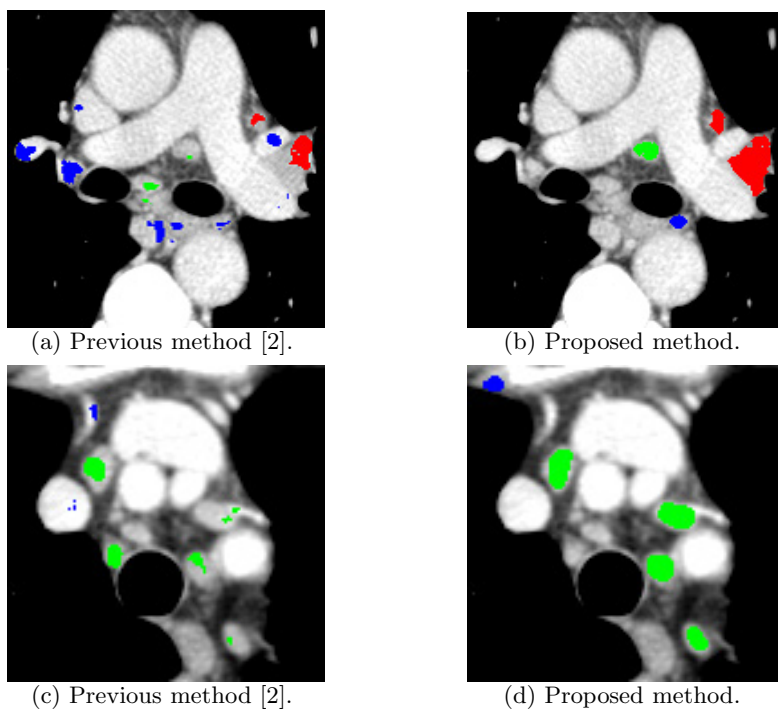
(a) Previous method [2].

(b) Proposed method.

(c) Previous method [2].

(d) Proposed method.

Fig. 4. Comparison of detection results obtained with the previous [2] and proposed methods (Red: lymph node metastases, Green: enlarged lymph nodes, and Blue: false positives). The proposed method extracted the lymph node regions more correctly.

4   Discussion

4.1   Accuracy Evaluation

The previous method detected only the central parts of the lymph nodes, and so it could not distinguish true lymph nodes from FPs. The proposed method significantly reduced the FP number since it extracted more precise lymph node regions by region growing which utilizes differences in shape between the lymph nodes and FPs (e.g., lymph nodes are round while FPs are long). However, the proposed method deleted a long shaped metastasis as shown in Fig. 6 (b) since the condition (iii) for the FP reduction was satisfied, though it was detected in the rough extraction step. Figure 7 shows the FROC analysis of the parameters used in the proposed method. The smaller the T_L (threshold to the Laplacian filter output) was, the smaller the FP number was, though the sensitivity decreased significantly. The parameter T_S (threshold to the diameter of lymph node candidates) controls the occurrence of FPs while keeping the sensitivity high. When decreasing the parameter T_F (threshold to the ratio of max/min diameters), true lymph nodes were also eliminated as FPs since some lymph nodes show ellipsoidal shape.

(a) Original image.

(b) Proposed method.

(c) Original image.
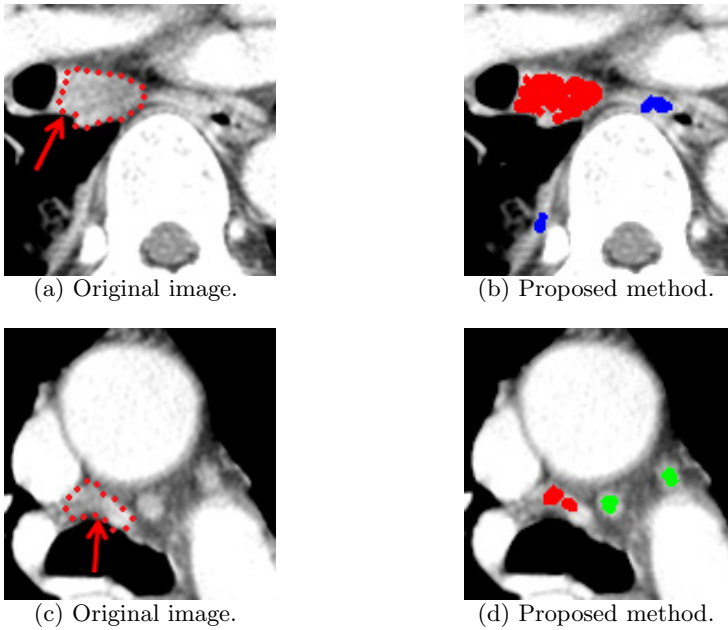
(d) Proposed method.

Fig. 5. Examples of true positive metastases detected by using both the proposed method and the radiologist (a,b) and only by using the proposed method (c,d). Red regions indicate lymph node metastases, green regions enlarged lymph nodes, and blue regions false positives.

(a) Example 1 (axial view).

(b) Example 2 (sagittal view).
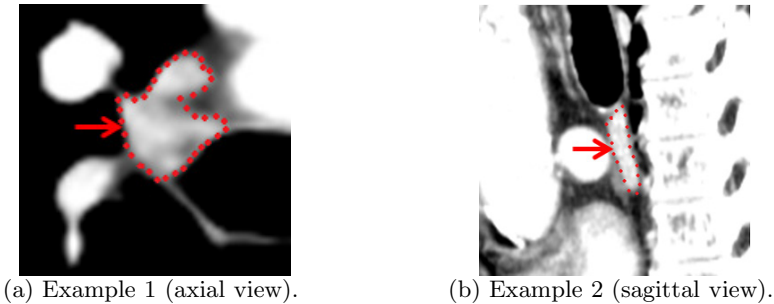
Fig. 6. Examples of false negatives that both the proposed method and the radiologist obtained

4.2   Clinical Evaluation

The proposed method detected lymph node metastases more accurately than the radiologist. Note that the radiologist is not a novice but an expert with about twenty years experience in image diagnosis. The sensitivity of radiologists is generally low because they usually point out only lymph nodes that can be
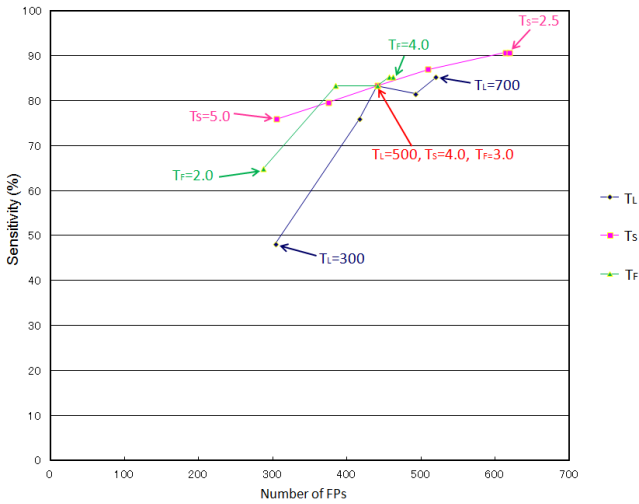
Fig. 7. FROC analysis of the parameters used in the proposed method

clearly recognized as metastasis lymph nodes in reports. This is suggested by the generally high specificity of radiologists. Although the proposed method must achieve a lower FP number for it to be put to clinical use, it is able to detect lymph node metastases that may be missed by a radiologist. The methodfs high sensitivity with reasonable specificity make it quite useful for application to surgical planning of lung cancer resection, in which it is absolutely essential not to miss any metastasis lymph nodes.

4.3   Visualization

In visualization of lymph nodes by virtual bronchoscopy, detected lymph nodes are well visualized through the semi-translucent bronchial wall. Note that the accuracy of assigning the number of stations has not been clinically evaluated yet. However, identification of the number of lymph node stations will be useful for both TBNA planning assistance and intra-operative guidance.

Table 1. Sensitivity and specificity of the previous [2] and proposed methods for the lymph nodes whose diameter is larger than 10 mm, including normal but enlarged lymph nodes

	Previous method[2]	Proposed method
Sensitivity	91.4 % (85/93)	89.2 % (83/93)
Specificity	3.3 % (85/2607)	16.8 % (83/493)
Num. of FPs/case	60.0 (2522/42)	9.8 (410/42)

Table 2. Comparison of the detection accuracy of lymph node metastases for the previous method [2], the proposed method, and image diagnosis by a radiologist. Note that normal enlarged lymph nodes are counted as FPs.

	Previous method[2]	Proposed method	Diagnosis of radiologist
Sensitivity	92.9 % (26/28)	92.9 % (26/28)	57.1 % (16/28)
Specificity	1.0 % (26/2607)	5.3 % (26/493)	94.1 % (16/17)
Num. of FPs/case	61.5 (2581/42)	11.1 (467/42)	0.02 (1/42)
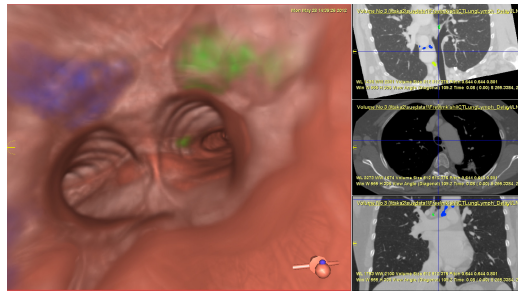
(a) Example 1 (near the carina).
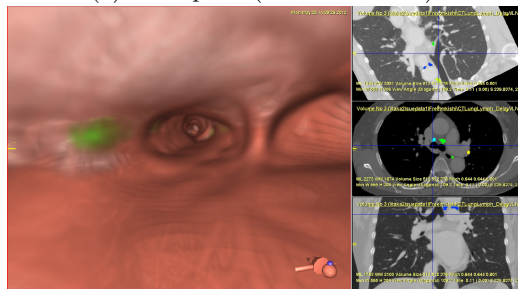
(b) Example 2 (near the second bifurcation).

Fig. 8. Examples of mediastinal lymph nodes visualized by virtual bronchoscopy. Different colors are assigned for each number of lymph node stations.

5   Conclusion

We proposed a method for automated detection of mediastinal lymph nodes from contrasted 3D chest CT images. The method uses local intensity structure analysis for rough extraction and region growing-based FP reduction for precise detection. In a lymph node detection experiment conducted using 42 pairs of contrasted CT images, the sensitivity and specificity for evaluating the methodfs accuracy were found to be respectively 89.2% and 16.8%, while those for a previous method were 91.4% and 3.3%. The values for evaluating the method clinically were 92.9 % and 5.3 %, while those for a radiologist were 57.1 % and 94.1 %. Visualizing detected lymph nodes by virtual bronchoscopy enables identification

of lymph nodes, which will be useful for TBNA guidance. The method has great promise for helping to prevent the risk of overlooking lymph node metastases, which will contribute to better and more effective lung cancer treatment.

References

1. Pass, H.I.: Mediastinal Staging 2005: Pictures, Scopes, and Scalpels. Seminars in Oncology 32, 269–278 (2005)
2. Feuerstein, M., Glocker, B., Kitasaka, T., Nakamura, Y., Iwano, S., Mori, K.: Mediastinal Atlas Creation from 3-D Chest Computed Tomography Images: Application to Automated Detection and Station Mapping of Lymph Nodes. Medical Image Analysis 16(1), 63–74 (2012)
3. Feulner, J., Kevin Zhou, S., Hammon, M., Hornegger, J., Comaniciu, D.: Segmentation Based Features for Lymph Node Detection from 3-D Chest CT. In: Suzuki, K., Wang, F., Shen, D., Yan, P. (eds.) MLMI 2011. LNCS, vol. 7009, pp. 91–99. Springer, Heidelberg (2011)
4. Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense image registration through MRFs and efficient linear programming. Medical Image Analysis 12(6), 731–741 (2008)
5. Feuerstein, M., Kitasaka, T., Mori, K.: Automated Anatomical Likelihood Driven Extraction and Branching Detection of Aortic Arch in 3-D Chest CT. In: The Second International Workshop on Pulmonary Image Analysis, MICCAI 2009, pp. 49–60 (2009)
6. Sato, Y., Westin, C.F., Bhalerao, A., Nakagima, S., Shiraga, N., Tamura, S., Kikinis, R.: Tissue Classification Based on 3D Local Intensity Structures for Volume Rendering. IEEE Transactions on Visualization and Computer Graphics 6(2), 160–180 (2000)
7. Nagao, M., Matsuyama, T.: Edge preserving smoothing. Computer Graphics and Image Processing 9(4), 394–407 (1979)

Author Index