Proceedings of

amiarcs 08

http://campar.in.tum.de/AMIARCS08/WebHome

Preface

These proceedings cover the AMI/ARCS one-day satellite workshop of MICCAI 2008. This workshop continued the tradition of Augmented Reality in Computer Aided Surgery (ARCS) 2003 and Workshop on Augmented environments for Medical Imaging and Computer-aided Surgery (AMI ARCS) 2004 and 2006. The workshop is a forum for researchers involved in all aspects of augmented environments for medical imaging. Augmented environments aim to provide the physician with enhanced perception of the patient either by fusing various image modalities or by presenting medical image data directly on the physician's view, establishing a direct relation between the image and the patient.

The workshop was divided into three sessions on applications, accuracy & assessment and display & visualisation. There was also a poster session describing interesting and varied aspects of research in this field. There were two invited speakers. Professor Yoshinobu Sato from the Division of Image Analysis, Osaka University, Japan, gave a talk entitled *Deformation for endoscopic surgical navigation* about his work in augmented reality for thoracic and abdominal surgery including clinical experience with his system. Dr Mirna Lerotic from the Institute for Bioemdical Engineering, Imperial College, London, UK, spoke about her work in augmented reality depth perception in a talk titled it Augmented reality for minimally invasive surgery.

There was a particlar focus on the path to clinical implementation and any barriers that might exist to implementation of systems in the operating room. A panel discussion on the future of clinical AR was held. This included two of the most prolific research directors in the field, Nassir Navab and Terry Peters. There was valuable clinical input from orthopaedic surgeon Dr. Sandro Michael Heining and the industry perspective on the route to market was given by Daniel Ruijters from Philips Healthcare.

AMI/ARCS 2008 brought together clinicians and technical researchers from both industry and academia with interests in computer science, electrical engineering, physics, and clinical medicine to present the state-of-the-art developments in this ever-growing research area. We are delighted at the ongoing high level of interest in AMI/ARCS, with a steady number of participants and the continued high quality of the submissions. We are also pleased to see that there is a trend in the research from proof-of-principle demonstrators towards real clinical applications that will benefit patients in the foreseeble future.

September 2008

Dr Philip "Eddie" Edwards
Dr Tobias Sielhorst
Prof Kensaku Mori
Workshop Chairs
AMI/ARCS 2008

Organization

AMI/ARCS 2008 is organized jointly by the Imperial College, London, UK, the Technical University of Munich, Germany and Nagoya University, Japan. It was held in the Helen and Martin Kimmel Center, New York University as a satellite workshop of MICCAI 2008.

Program Committee

Wolfgang Birkfellner
Stephane Nicolau
Joerg Traub
Nassir Navab
Terry Peters
Jannick Rolland
Adrien Bartoli
Marie-Odile Berger
Rudy Lapeer
George Stetten
Ali Khamene
Sandro Heining

Table of Contents

Augmented Environments for Medical Imaging and Augmented Reality in Computer-aided Surgery

A Training Suite for Image Overlay and Other Needle Insertion Techniques

Gregory S. Fischer[1], Iulian Iordachita[1], Paweena U-Thainual[2], John A. Carrino[3], and Gabor Fichtinger[1, 2]

[1] Center for Computer Integrated Surgery, Johns Hopkins University, Baltimore, MD, USA
[2] School of Computing, Queen's University, Kingston, ON, Canada
[3] Department of Radiology, Johns Hopkins Hospital, Baltimore, MD, USA

Abstract. In order to develop accurate and effective augmented reality (AR) systems used in MR and CT guided needle placement procedures, a comparative validation environment is necessary. Clinical equipment is prohibitively expensive and often inadequate for precise measurement. Therefore, we have developed a laboratory validation and training system for measuring operator performance using different assistance techniques. Electromagnetically tracked needles are registered with the preoperative plan to measure placement accuracy and the insertion path. The validation system provides an independent measure of accuracy that can be applied to various methods of assistance ranging from augmented reality guidance methods to tracked navigation systems and autonomous robots. In preliminary studies, this validation system is used to evaluate the performance of the image overlay, bi-plane laser guide, and traditional freehand techniques. Perk Station, an inexpensive, simple and easily reproducible surgical navigation workstation for laboratory practice incorporating all the above mentioned functions in a "self-contained" unit is introduced.

Keywords: validation system, augmented reality, image overlay, mri, needle placement, percutaneous procedures, image guided surgery,

1 Introduction

In recent years, numerous surgical methods have been developed for needle-based surgery. Image-guided percutaneous needle-based surgery has become part of routine clinical practice in performing procedures such as biopsies, injections and therapeutic implants. Contrary to casual observation, needle-based surgery can be an exceedingly complex intervention. Translation and rotation motions, as well as bending and insertion forces can be combined for delicate needle control in needle-based surgery. The workspace and the means for guiding the surgical device, however are extremely limited. Last but not least, detecting and recovering from errors increases the risk of these otherwise appealing outpatient procedures.

Trainees usually perform needle interventions under the supervision of a senior physician. This is a slow and inherently subjective training process that lacks objective, quantitative assessment of the surgical skill and performance. A comparative validation environment is necessary for an efficacious analysis of CT/MRI guided assistance techniques to be used in needle placement procedures. Clinical equipment is prohibitively expensive and often inadequate for precise validation. Precise measurement of placement accuracy by MRI is greatly limited by paramagnetic needle artifact and lack of distinct small targets. Scanner time cost can exceed $500/hour making statistically significant trials impractical. To address these

issues, we have developed a laboratory validation system for measuring operator performance of different assistance techniques and furthermore we are developing the *Perk Station*, an inexpensive, simple and easily reproducible surgical navigation workstation for laboratory practice with non-biohazardous specimens.

The validation system can be applied to various methods of assistance ranging from augmented reality guidance methods to tracked navigation systems. Preliminary accuracy assessment of our MR image overlay system has been performed, but the excessive cost of scanner time has thwarted a large-scale study of the accuracy of this system. Therefore, an off-line validation system has been created in order to study needle placement accuracy; in particular we look at the accuracy of the image overlay and compare it to that of other insertion guidance methods. This system will also provide a means to study the trajectory and gestures [1] throughout the insertion procedure in addition to the endpoint accuracy. The study of hand gestures for each of these methods will provide useful information that can be used to help minimize the number of re-insertion attempts needed, as each re-insertion causes significant discomfort to the patient. This system provides a less resource exhaustive and more accurate means by which to validate needle insertion procedures.

In this paper, we describe the needle insertion validation system shown in Fig. 1, its use for comparative analysis of the virtual image overlay, the bi-plane laser guide and unassisted freehand techniques. We also introduce the *Perk Station,* a "self-contained" system designed as a clinical training and evaluation system.

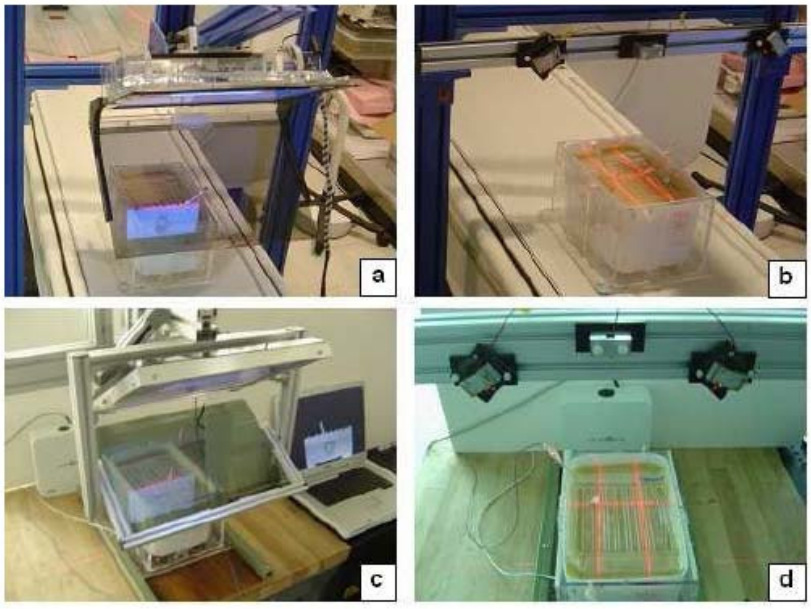


Figure 1. Image overlay (a, c) and bi-plane laser guide (b, d) AR needle placement systems with spine phantom. Feasibility trials in the MRI scanner (a, b) and the validation system with EM tracked needles in the laboratory (c, d).

2 System Description

The image overlay consists of a flat display and a half-silvered mirror mounted on a gantry as seen in Figure 1 (a,c). After calibration, when the physician looks at the patient through the mirror, the CT/MR image appears to be floating inside the body with the correct size and position as if the physician had 2D 'X-ray vision' regardless of view angle without any external tracking. Prior to needle insertion the image is transferred directly in DICOM format to the planning and control software running on a stand-alone laptop where we mark the target and entry points, draw a visual guide along the trajectory of insertion, mark the depth of insertion and push this image onto the overlay display [2]. The physician inserts the needle using the traditional workflow, while the overlay provides in-situ anatomical and guidance information in the image plane. The laser guide uses two laser planes; one transverse plane and one oblique sagittal plane as seen in Fig. 1 (b,d). The intersection of these two laser planes marks the needle insertion path. For convenience, a second oblique sagittal laser can be added to support bilateral interventions [3]. In tracked navigation [4], the planned needle path can be superimposed in orthogonal planes, in oblique plane including the needle, or in transparent volumes

2.1 Validation System

Electromagnetic (EM) tracking (Aurora, Northern Digital, Waterloo, Ontario) is utilized to provide the position of the tip and orientation of the shaft of an instrumented needle as described in [5]. All necessary components must be registered with one another in order to track the needle with respect to the preoperative plan generated on the MR/CT images. The components of the system include: the Aurora EM Tracker, a tracked needle, the tracked phantom, the MR/CT images used for pre-operative planning and the AR guidance system (e.g. image overlay or laser guide). The system is shown in Fig. 1.

2.2 Phantom Design

A human cadaver lumbar spine phantom was designed to mimic the anatomy of a patient and aid in the process of registration. Lumbar vertebrae and simulated intravertebral discs are placed in proper alignment and are embedded into a layered tissue mimicking gel (SimTest, Corbin, White City, OR) of two different densities emulating fat and muscle tissue. The gel phantom with lumbar spine is placed into an acrylic enclosure which was accurately laser-cut with 28 different pivot points spread over four sides for rigid-body registration. Stereotactic fiducial markers (MR-Spots, Beekley, Bristoll, CT) were placed on the phantom in precisely positioned laser-cut slots. The markers were placed in a 'Z' shape pattern on three sides allowing for automatic registration between anatomical images and the phantom. Modularity allows other anatomical phantom to be placed in the enclosure.

2.3 MR Image Registration

In order to register preoperatively obtained MR or CT images (and their respective preoperative plans) to their corresponding physical space, techniques similar to those described in [6] are used (Fig. 2).

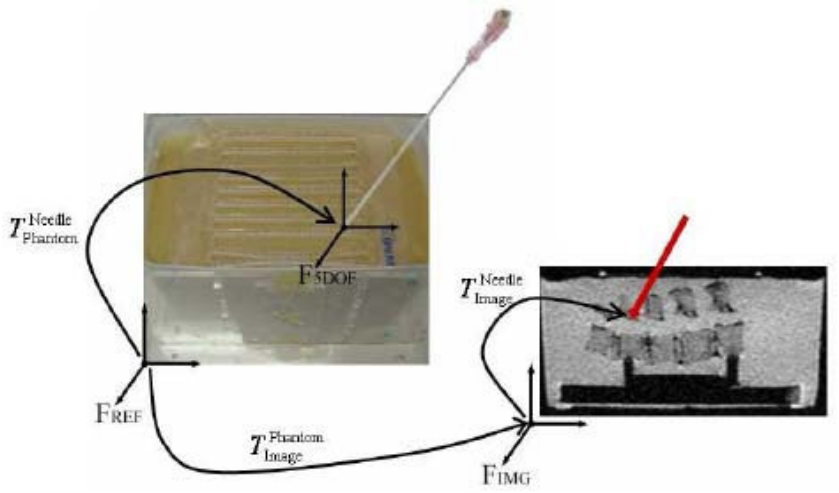


Figure 2. Frame transformations for the registration process shown on the spine phantom and its corresponding MR images. The tracked needle is represented in the original image space where the preoperative plan was made.

Axial images are taken near the center of the phantom; the locations of other images of the phantom are known with respect to this reference. The central image is used for registration, here the nine fiducial markers arranged in three Z-shaped patters are segmented by applying an adaptive threshold and morphological operations to the image. The centroid of each marker was then found and the position of each marker with respect to the DICOM image was recorded into a set of nine points. After the nine distinct points were identified, the transformation from the scanner's image space to the phantom's coordinate system was computed. The RMS error incurred in the image to phantom space registration for a typical MR image was 1.26mm.

2.4 Electromagnetic Tracker Registration

The NDI Aurora EM tracking system is used to localize an instrumented needle with respect to the phantom. A six degree-of-freedom (DOF) reference tool is fixed to the phantom and a calibrated pointer tool is used for rigid-body registration of the phantom to the tracker. Data was obtained by pivoting about 24 pre-defined divot points with the pointer. These points were used for registration between phantom coordinate system and that of the EM tracker by finding the transformation which aligns the known point locations obtained from the mechanical design specifications with the collected data points. The RMS error incurred in the rigid-body registration was 0.93mm. Once both steps in registration are complete, an instrumented needle may be tracked as it maneuvers along a planned path within the phantom. To maximize the system's accuracy, future efforts will include distortion mapping and error compensation similar to that described in [7].

2.5 Perk Station

The *Perk Station* comprises image overlay (Fig. 1 a,c), laser guide (Fig. 1 b), and standard tracked freehand navigation (Fig. 1 d), in a single suite. The general system concept is shown in Fig. 3. A detailed view of the *Perk Station* is shown in Fig. 4.

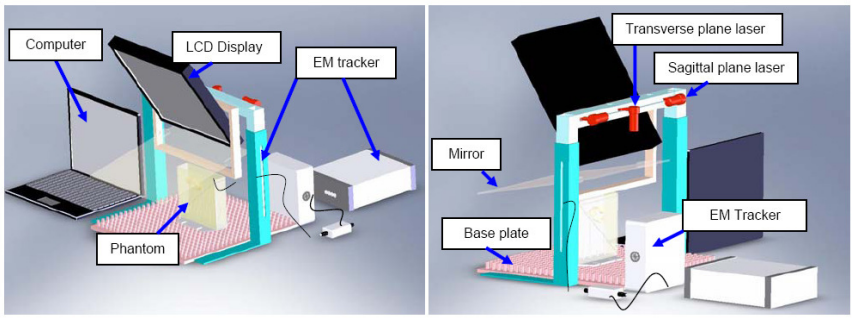


Figure 3: CAD design of the Perk Station: image overlay (left) and laser guide (right) shown with EM tracked navigation system.

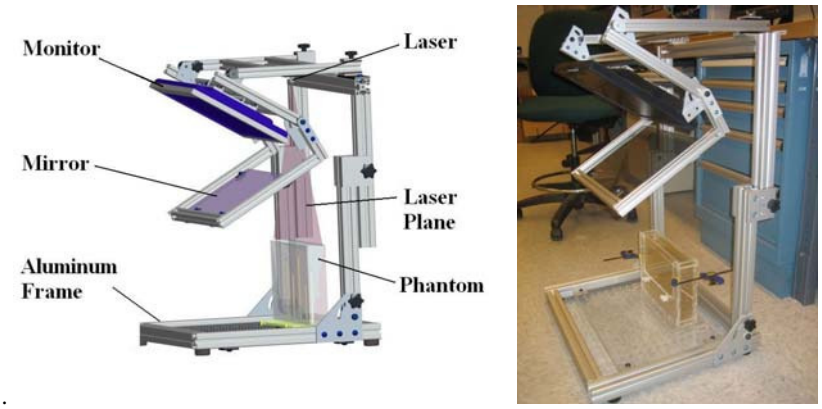


Figure 4: Perk Station with Image Overlay unit. CAD model (left) and physical realization (right) of the current design.

In all cases, the needle is tracked in real-time during the course of the insertion, not only the final tip position. The Aurora electromagnetic tracking system provides the tip position and axis orientation of the needle. The image overlay is mounted on one side and the laser guide and tracked navigation system on the opposite side. The user can swap between guidance techniques by turning the system around. Traditional freehand interventions can be evaluated by only using the transverse laser plane, functionally similar to the ones on the face of the MR scanner. The extruded aluminum frame is sufficiently strong to hold the weight of all components, yet the system is still sufficiently lightweight to be portable in a suitcase.

Another important part of the system is the "real-time" nature of the phantom shown in Fig. 5. Geometrical or anatomical phantoms are housed in an interchangeable rigid box (inside box). A reusable external housing (outside box) is equipped with external markers (stereotactic fiducials and EM tracking coils), and can be easily realigned under the overlay. The phantom and needle are tracked with respect to each other and registered to the MR/CT images. Using such phantoms, we can acquire MRI images for targeting, but then perform the needle placement in the laboratory. This arrangement makes it possible to perform hundreds of needle placements without holding up the MRI scanner, while the accuracy of electromagnetic needle tracking significantly surpasses the usual accuracy of MRI-based evaluation.

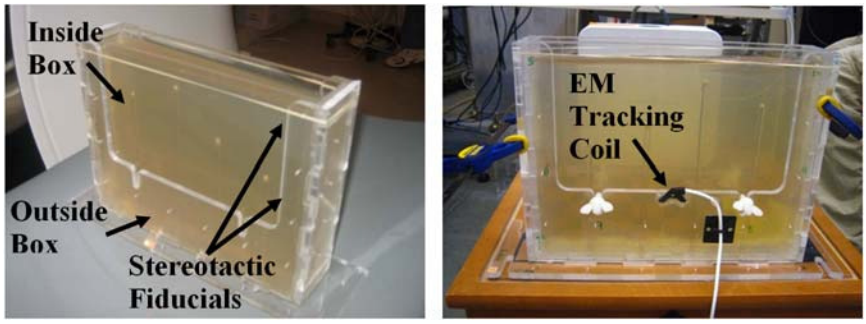


Figure 5: The "real-time" phantom.

3 Experimental Methods

Prior to beginning trials, numerous needle paths were created in the EasySlice planning software that the authors have developed and is described in [2]. The software stores the insertion and target points for each planned path as well as the angle of insertion needed to accurately reach the desired target. For each of the three needle insertion methods (image overlay, bi-plane laser guide and freehand interventions) presented, subjects were randomly assigned three different paths in three different axial MR slices. The entire insertion attempt was recorded with the tracking software. The software then provides insertion and target point error, both in and out of the image plane. Needle axis orientation error is also computed. Simple forms of gesture tracking are now provided, including distances from the trajectory during insertion and the number of re-insertion attempts.

4 Results

To demonstrate workflow, four needle insertions were performed with each technique in a clinical MRI environment. As expected, accuracy could not be assessed due to large artifacts as shown in Fig. 6(a). In the validation testbed, the measured needle trajectories were graphically overlaid on the plan and targeting MR image as shown in Fig. 6(b). Twenty insertions were performed with each technique. Position and orientation errors were measured. Initial analysis showed that the results correlate with direct validation performed using fluoroscopy described in [3]. The

image overlay's mean error in the image plane was 1.4mm and 2.5° with standard deviations of 0.5mm and 1.9° respectively. The laser guide's average error was 1.8mm and 2.0° (1.2mm and 1.8° standard deviation), and freehand produced average errors of 2.0mm and 5.2° (1.4mm and 2.3° standard deviation).

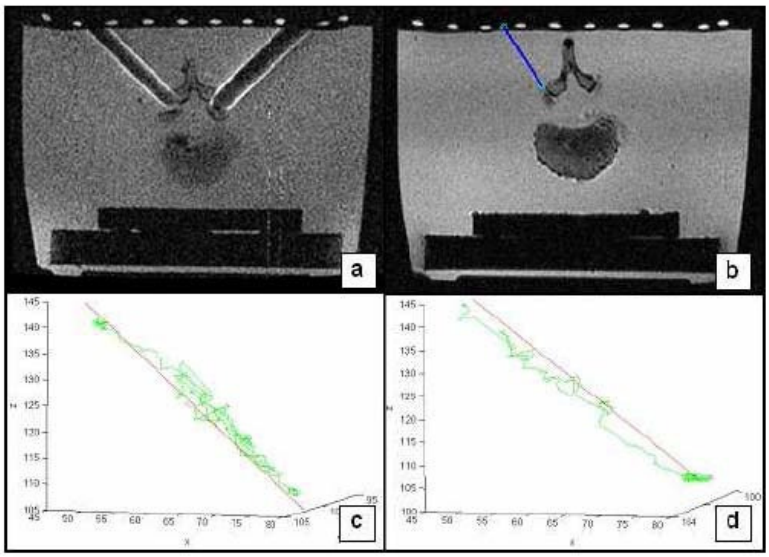


Figure 5. Typical results from an MR image (a) and a tracked path (b). Logged trajectory showing multiple corrections (c) and direct path (d).

5 Discussion and Conclusion

Initial assessments of the image overlay, laser guide, and freehand needle insertions were performed with the validation system. Experiments with experienced radiologists are currently underway. Future experiments will provide independent, large scale accuracy assessment of needle insertion procedures using commercial surgical navigation systems, image overlay, laser guidance, and traditional techniques. The goal is to quantitatively compare placement accuracy, consistency, and other important characteristics such as the needle trajectories throughout the entire placement procedure.

In typical needle placement procedures, the interventionalist will often probe the patient's anatomy until the desired target is reached. This probing action can result in a great deal of discomfort to the patient as well as significant bruising to the area. Analysis of the needle trajectory can provide information about the number of insertion and repositioning attempts that were made during an intervention. Fig. 5(c) illustrates a trajectory that resulted from repeated reinsertions and Fig. 5(d) shows an insertion with minimal repositioning attempts. This information enables researchers to study the systems' ability to minimize discomfort to the patient during the procedure.

The *Perk Station* is designed to be a replicable and adaptable tool for teaching computer-assisted surgery at all levels. To promote transferability, the complete design of the Perk Station, including hardware blueprints, phantoms blueprints, and software source code will be made publicly available as open source. Simple design and low costs allows interested parties to replicate the hardware and install the software. CT/MRI data and pre-made surgical plans will also be provided, so that users can operate the Perk Station without having access to medical imaging facilities. It is a small, portable, and light weight, and fits in a suitcase when disassembled. The apparent simplicity of the *Perk Station* should not underestimate its potential in teaching and training medical professionals, particularly medical students and residents. There is a general misperception and under-appreciation among the public of the skills required for needle based surgeries. In reality, trainees gravitate to learning centers where procedural skills are taught. There is popular trend toward minimization the steep learning curve associated with surgical interventions through the use of simulators. Static or declining reimbursements have driven the need for economical solutions: training systems of with accuracy, efficiency, simplicity, and low cost. The *Perk Station* promises to fit in these trends eminently.

Acknowledgements. This work has supported by Queen's University Teaching and Learning Enhancement Grant, U.S. National Institutes of Health 1R01CA118371-01A2, and the National Alliance for Medical Image Computing (NAMIC), funded by the National Institutes of Health through the NIH Roadmap for Medical Research, U54 EB005149.

References

1. Lin HC, Shafran I, Yuh D, Hager GD. Towards automatic skill evaluation: detection and segmentation of robot-assisted surgical motions. Comput Aided Surg. 2006; 11(5):, 220–230.
2. Fischer GS, Deguet A, Schlattman D, Taylor RH, Fayad L, Zinreich SJ, Fichtinger G. MRI image overlay: applications to arthrography needle insertion. Studies in health technology and informatics - Medicine Meets Virtual Reality 14, 2006; 119:150–155.
3. Fischer GS, Wamsley C, Zinreich SJ, Fichtinger G. Laser-Assisted MRI-Guided Needle Insertion and Comparison of Techniques. Int Soc for Comp Asst Orthopaedic Surg, Jun 2006.
4. Image-Guided Interventions: Technology and Applications, Terry Peters, Kevin Cleary, Springer, 2008.
5. Wood BJ, Zhang H, Durrani A, Glossop N, Ranjan S, Lindisch D, Levy E, Banovac F, Borgert J, Krueger S, Kruecker J, Viswanathan A, Cleary C. Navigation with Electromagnetic Tracking for Interventional Radiology Procedures: A Feasibility Study. J Vasc Interv Radiol. 2005; 16:493-505.
6. Lee S, Fichtinger G, Chirikjian GS. Novel algorithms for robust registration of fiducials in CT and MRI. J Med Phys 2002;29(8):1881-1891.
7. Fischer GS, Taylor RH. Electromagnetic Tracker Measurement Error Simulation and Tool Design, Med Img Comp & Comp Asst Int. 2005; LNCS 3750:73-80.

A 3D Stereo System to Assist Surgical Treatment of Prostate Cancer

Dongbin Chen, Lin Mei, Ara Darzi, and Philip Edwards

Dept. of Biosurgery and Surgical Technology, Imperial College, London, UK
dchen1,l.mei,a.darzi,eddie.edwards@imperial.ac.uk *

Abstract. This paper presents our work to establish the feasibility and utility of a system for guiding robotic removal of the prostate (radical prostatectomy). The guidance is based on preoperative imaging aligned to the patient on the operating table and presented to the surgeon as graphics overlaid on the stereo laparoscopic view through the da Vinci robot. The results of preoperative and intraoperative data collection and processing are presented, along with stereo calibration and visualisation on the surgical scenes. We have gathered data on 8 patients for retrospective evaluation. We have identified the relevant anatomy that should be modelled: the prostate itself, urethra, surrounding nerve bundle, seminal vesicles and the pelvic bone for registration. Stereo visualisation provides good depth perception, which we will present at the workshop. At present registration is by direct alignment to the endoscope view, but we plan to incorporate transrectal ultrasound to compensate for soft tissue motion. There will be full clinical evaluation on 20 patients in the next phase of this project. From the outset we have considered the path to clinical implementation and guidelines are presented that could be followed by any researchers aiming to provide augmented reality surgical guidance.

1 Introduction

1.1 Clinical Need

Prostate cancer is increasing in an aging population and is now the most common form of cancer among men in the UK. Robot-assisted prostatectomy is becoming an accepted form of surgical treatment [1]. There are significant rates of complication, however, from both robotic and open radical prostatectomy. These include impotence, incontinence and recurrence of disease due to incomplete resection. We propose that complication rates can be reduced for robotic procedures using augmented reality (AR) guidance. Overlay of correctly aligned 3D anatomy and pathology derived from preoperative imaging should help the

* We would like to thank Cancer Research UK for funding this project (C24520/A8087). We are also grateful to many other members of the Department of Computing and the Department of Biosurgery and Surgical Technology at Imperial College as well as the radiology and theatre staff at St. Mary's Hospital, London.

surgeon to achieve full extraction of the diseased tissue while preserving as much as possible of the surrounding neurovascular bundle, leading to better outcomes for the patient.

1.2 Previous Work

Image guidance is now an accepted tool in neurosurgery, ENT, maxillofacial surgery and orthopaedics [2, 3]. Visualisation usually shows the location of a pointer or tool in orthogonal cuts through the preoperative dataset, but AR solutions showing virtual structures overlaid on an optical or camera view of the patient have also been proposed [4–6].

Within the abdomen image guidance has been proposed to aid liver surgery [7, 8], but though AR laparoscopy has been suggested [9], only a single clinical demonstration of AR guidance for adrenalectomy has so far been reported [10].

There are a number of technical hurdles that must be overcome to achieve AR guidance. The optics of the stereoendoscope must be calibrated. The preoperative scan must be registered to the physical position of the patient relative to the endoscope coordinate system. If there is tissue deformation, this registration is non-rigid. Finally, the aligned model must be presented to the surgeon in real time using visualisation that enables correct 3D perception of structures.

In this work we aim to follow a clear path that should lead to a clinically useful system. This follows a three stage approach — retrospective evaluation, intraoperative evaluation and initial clinical use. In all phases the involvement of clinicians is key and the gradual introduction of equipment into the operating room will ease the acceptance the system. We are currently approaching the end of the first phase of this project, where preoperative imaging and intraoperative stereo video data are gathered and evaluated. We present the hardware configuration of the system and describe 3D model construction from preoperative data, stereo optical calibration and our implementation of AR visualisation. Results are given for calibration as well as evaluation on a phantom and 8 patients.

2 Methods

2.1 Hardware Configuration and System Integration

Hardware configuration of the system is shown in Figure 1. The da Vinci robot has a master and slave configuration. The stereoendoscope video images are captured using a multi-input framegrabber (the Active Silicon LFG4) with input from the camera controller on the da Vinci stack. The system software runs on a 64-bit linux workstation (HP XW9400) with dual graphics output (2 x NVidia Quadro FX1400). The graphics and video are merged by two video mixers (Edirol LV4) using the chroma-key function. We decided to use video mixers rather than merging the graphics on the workstation since this ensures there is no lag introduced and means that the normal surgical view is not dependent on the function of the workstation. An optical tracker can be incorporated (the

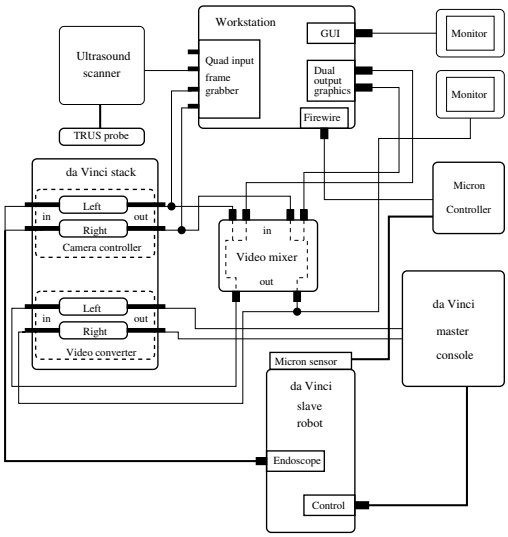


Fig. 1. The system configuration. The guidance is provided from a workstation which takes as input the stereo video from the da Vinci robot along with tracking and registration data from the MicronTracker^{TM} and ultrasound scanner and produces aligned graphics. AR is provided using two video mixers to avoid lag in display of the real data

MicronTracker2 from Claron Technologies). Though not yet incorporated, we aim to include an ultrasound scanned for tracking of soft tissue motion during surgery.

The proposed system provides 3D visualisation of graphics overlaid on the surgical scene through the robot's stereoendoscope. We will display preoperative image information showing the location of anatomical and pathological structures accurately aligned to the patient. The 3D nature of the overlaid virtual features will enable the surgeon to view structures such as the vas deferens, seminal vesicles and the neurovascular bundle beneath the operative surface, effectively providing "X-ray eyes".

2.2 Preoperative Imaging Model

As the first step in deciding on an imaging model, we identified the relevant anatomy in consultation with clinicians. For guidance of surgery several features are of interest. These include the prostate itself, the urethra, rectum, seminal vesicles and the surrounding nerve bundle. Most of these can be seen on MRI, though the nerves can be difficult to identify. In addition to the local anatomy we need to extract the bone of the pelvis for registration. This would ideally come from a CT scan registered to MRI, but we would like to build our model from MRI alone. This can be problematic, since bone does not enhance with

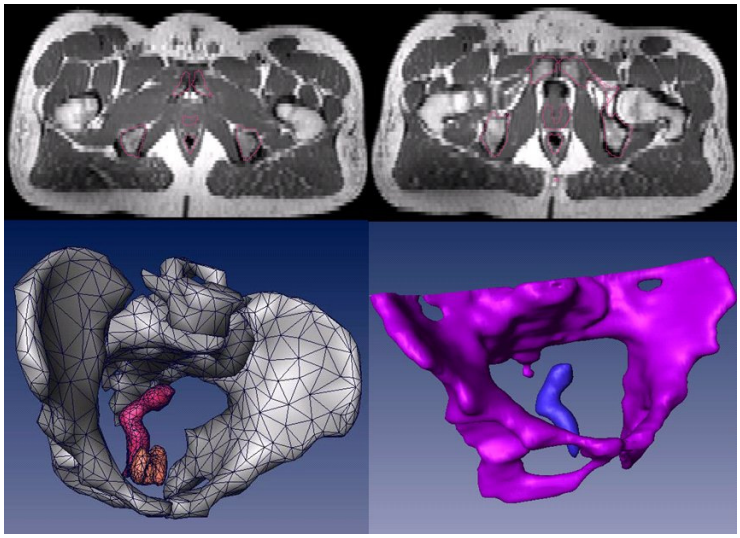


Fig. 2. Atlas segmentation. Overlaid on the patient MRI (top) we see the segmentation resulting from registration of the Atlas to the patient. A 3D rendering of the warped atlas (bottom left) and corresponding manual segmentation (bottom right) are shown.

MRI. We are examining whether registration to a template CT scan can provide bone segmentation from MRI.

For our initial group of patients, manual segmentation has been performed to create a 3D model. This is a laborious task. The advantage of expert manual segmentation is the high accuracy of results. We are working on automated atlas segmentation guided by a statistical shape model of the lower abdomen. To this end we are building a database of scans to ensure sufficiency of the training set and are developing techniques to establish that the shape model has stabilised. Comparing the results with expert segmentation, we achieve errors of the order of 2 voxels.

We have used atlas segmentation to obtain 3D models from 12 patients respectively. Figure 2 shows a 3D model of prostate, pelvis and rectum segmented from a patient's scan data. Comparing the warped atlas to the manually segmented patient model, we can see that that there are some small differences. We are examining whether a final correction of the registration can be applied to improve the results further.

2.3 Stereo Camera Calibration

Stereo camera calibration is vital component for accurate overlay on the 3D surgical scene. We use Bouguet's Matlab camera calibration toolbox [11] to calibrate the stereo camera system. We first determine intrinsic calibration parameters to each endoscope. These parameters include internal parameters of the pinhole

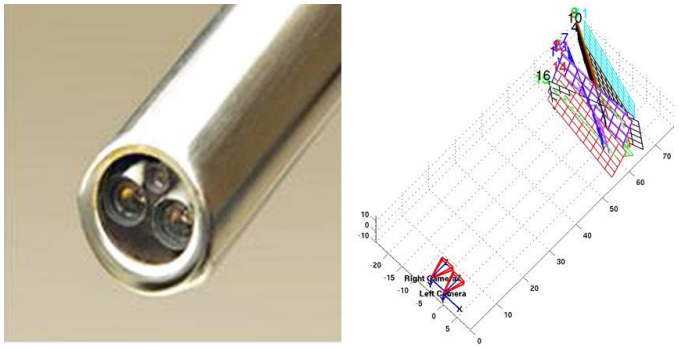


Fig. 3. The da Vinci stereo endoscope and an example of the resulting calibration from the Matlab toolbox

model plus parameters describing lens distortion. The position and orientation of one camera is established with respect to the other in the stereo phase of the calibration. Figure 3 gives the reconstruction of locations between the stereo camera and images captured for the calibration.

2.4 System for Augmented Reality Guidance and Retrospective Overlay

As well as providing a system for live overlay in the operating room, we have developed software for retrospective augmented reality. We can also control the position and orientation of the 3D models overlaid on recorded intraoperative video. The software interface displays a 3D surgical scene with overlay as a pair of stereo images, which can be viewed using a simple set of prism glasses. This retrospective overlay system enables experimentation with different rendering paradigms and allows us to examine which visualisations provide useful guidance and accurate depth perception. This can be done without taking up time in the operating theatre.

For live overlay, the workstation simply displays graphics on a blue background. The video mixers then merge this with the live endoscopic video using the blue color as the background colour for chroma-keying. Registration is currently achieved by manual alignment with the visible pelvic arch, but we are working on automated alignment to the stereo video and refinement using transrectal ultrasound.

3 Results

Figure 4 shows overlay on a prostate phantom. The overlay shows pelvis, prostate, urethra and seminal vesicles. Figure 5 shows an example of augmented reality on a real intraoperative patient view. We have collected such data on 8 patients and will present these at the workshop.

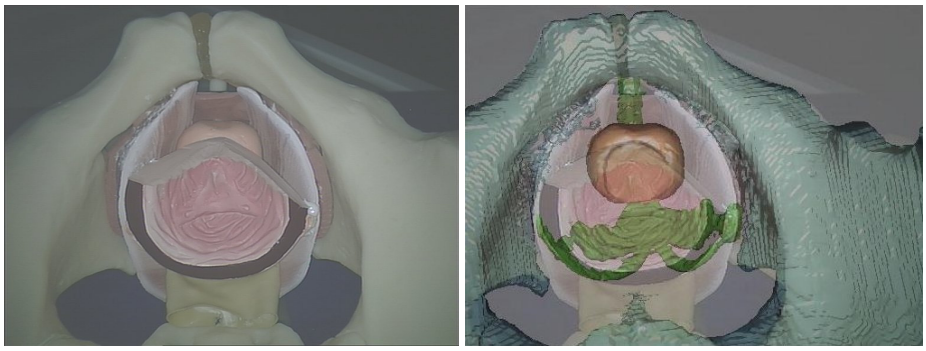


Fig. 4. Overlay on the anatomical phantom of the pelvis and prostate. The visible structures (e.g. pelvis) are seen to be correctly aligned and the underlying prostate and seminal vesicles are visible in the augmented view.

4 Guidelines and Principles for the Clinical Introduction of Augmented Reality

As with all systems that aim to provide augmented reality surgical guidance, our intention is to have a clinically usable system that provides benefits for the surgeon and patient. Our experience in developing such systems has led to a number of guidelines that may be useful to other researchers.

4.1 The Three Stage Approach

We propose the introduction of AR in the following three phases.

1. Retrospective evaluation
2. Intraoperative evaluation
3. Initial clinical introduction

In the first stage, information preoperative imaging and intraoperative video data is gathered. Evaluation can then be performed after the operation. Identification of critical points in the procedure that could benefit from guidance is possible at this point and many different visualisation methods can evaluated. The gathered data also enables testing of algorithms for registration. The result of this phase should be a protocol for both the software and the intraoperative workflow, identifying which anatomy or pathology should be shown and when.

In the second phase the system should be fully functional in-theatre. The surgeon can then view and evaluate the displays and provide vital feedback. At this point, however, the surgeon should only be evaluating the system and the guidance must not be used to influence surgical decision making. This enables the guidance methodology to be tested and developed in accordance with clinical requirements and should result in a stable system.

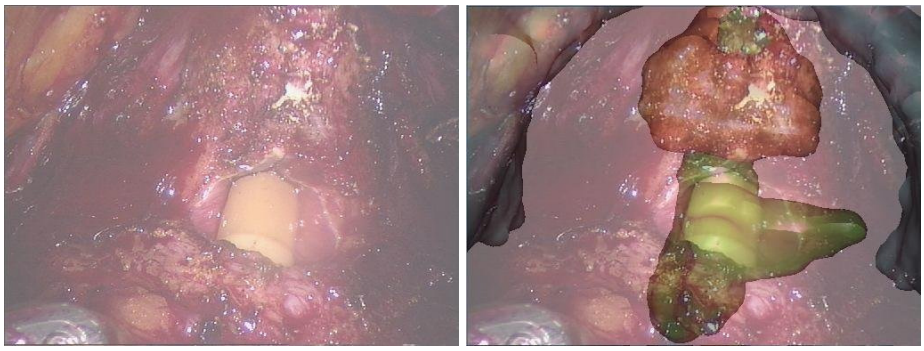


Fig. 5. An example overlay on a patient. Alignment is achieved using the pubic arch. The urethra is seen to match well (green). The position of the prostate before excision is shown in red.

The final phase is clinical evaluation where the surgeon can use the system to influence the procedure. Since the surgeon will be familiar with the system from the previous phase this reduces the risk of incorrect decision making, perhaps from inaccurate guidance or overestimation of the system accuracy.

Having these three clear phases give a simple and relatively safe process for the introduction to the operating theatre and can be useful when obtaining ethical approval.

4.2 General Principles for Clinical AR

As a result of our experience over many years, we have adopted the following principles that might be useful for other AR researchers to consider:

1. Ensure that there is a clinical need for guidance
2. Use the three stage approach (see above)
3. Use existing equipment wherever feasible
4. Introduce innovations one at a time (as far as possible)
5. Obtain and document feedback from clinicians
6. Minimise disruption to the normal clinical workflow
7. Ensure existing systems are available as a fallback
8. Do not allow clinical reliance on research systems until stage 3

Failure to follow these guidelines may lead to slow adoption of AR technology.

5 Discussion and Future Work

We have presented a system for augmented reality guidance of robotic prostatectomy. Though we are currently at the first phase of retrospective overlay,

the path to clinical implementation in the next two years is clear. We have described principles and guidelines for the adoption of AR into the clinical setting that other researchers should consider for their systems.

There are a number of technical developments in progress. We are automating the segmentation and preoperative model building from MRI. We would like to have a live optical calibration that can be used during the operation. The principal technical issue for any guidance system is registration. We are working on automated rigid alignment to the stereoendoscope view and updated alignment using transrectal ultrasound to follow soft tissue deformation. It is expected that the overall system will lead to reduced rates of incontinence, impotence and disease recurrence to the benefit of prostate cancer patients.

References

1. Kaul, S., Savera, A., Badani, K., Fumo, M., Bhandari, A., Menon, M.: Functional outcomes and oncological efficacy of vattikuti institute prostatectomy with veil of aphrodite nerve-sparing: an analysis of 154 consecutive patients. BJU Int. **97** (2006) 467–472
2. Aquilina, K., Edwards, P.J., Strong, A.J.: Principles and practice of image-guided neurosurgery. In Moore, M.A., ed.: Springer Specialist Surgical Series: Neurosurgery (In Press). Springer Verlag (2003)
3. Wadley, J.P., Dorward, N.L., Breeuwer, M., Gerritsen, F.A., Kitchen, N.D., Thomas, D.G.T.: Neuronavigation in 210 cases: Further development of applications and full integration into contemporary neurosurgical practice. In Lemke, H., Vannier, M., Inamura, K., Farman, A., eds.: Proceedings of Computer-Assisted Radiology and Surgery 1998, Elsevier, Amsterdam (1998) 635–640
4. Edwards, P.J., King, A.P., Maurer, Jr., C.R., de Cunha, D.A., Hawkes, D.J., Hill, D.L.G., Gaston, R.P., Fenlon, M.R., Jusczyzck, A., Strong, A.J., Chandler, C.L., Gleeson, M.J.: Design and evaluation of a system for microscope-assisted guided interventions (MAGI). IEEE Trans. Med. Imaging **19**(11) (2000) 1082–1093
5. Birkfellner, W., Figl, M., Matula, C., Hummuel, J., Hanel, R., Imhof, H., Wanschitz, F., Wagner, A., Watzinger, F., Bergmann, H.: Computer-enhanced stereoscopic vision in a head-mounted operating binocular. Phys. Med. Biol. **48** (2003) N49–N57
6. Rosenthal, M., State, A., Lee, J., Hirota, G., Ackerman, J., Keller, K., Pisano, E.D., Jiroutek, M., Muller, K., Fuchs, H.: Augmented reality guidance for needle biopsies: An initial randomized, controlled trial in phantoms. Med. Image Anal. **6** (2002) 313–320
7. Herline, A.J., Herring, J.L., Stefansic, J.D., Chapman, W.C., Galloway, R.L., Dawant, B.M.: Surface registration for use in interactive, image-guided liver surgery. Comput. Aided Surg. **5**(1) (2000) 11–17
8. Penney, G.P., Blackall, J.M., Hamady, M.S., Sabharwal, T., Adam, A., Hawkes, D.J.: Registration of freehand 3D ultrasound and magnetic resonance liver images. Med. Image Anal. **8** (2004) 81–91
9. Fuchs, H., Livingston, M.A., Raskar, R., Colucci, D., Keller, K., State, A., Crawford, J.R., Rademacher, P., Drake, S.H., Meyer, A.A.: Augmented reality visualization for laparoscopic surgery. In: Proc. Medical Image Computation and Computer-Assisted Intervention. (1998) 934–943

10. Marescaux, J., Rubino, F., Arenas, M., Mutter, D., Soler, L.: Augmented-reality assisted laparoscopic adrenalectomy. J. Am. Med. Assoc. **292**(18) (2004) 2214–2215
11. Bouguet, P.: A camera calibration toolbox for matlab

Where Cardiac Surgery Meets Virtual Reality: Pains and Gains of Clinical Translation

C.A. Linte[1,2], J. Moore[1], A. Wiles[1,3], J. Lo[1,2], D. Pace[1,2], C. Wedlake[1], D. Bainbridge[1,2], G.M. Guiraudon[1,2], D.L. Jones[1,2], and T.M. Peters[1,2,3]

[1] Robarts Research Institute - London ON Canada
[2] University of Western Ontario, London ON Canada
[3] Canadian Surgical Technologies and Advanced Robotics - London ON Canada
{clinte,tpeters}@imaging.robarts.ca

Abstract. Off pump intracardiac surgery is an area of research with many unique challenges. Surgical targets are in constant rapid motion in a blood filled environment that prevents direct line of sight guidance. The highly restrictive workspace requires compact yet robust tools for proper delivering therapy. In this paper we describe a comprehensive system currently under development for the treatment of conditions including mitral valve replacement, atrial septal defect repair and ablation therapies for atrial fibrillation. We describe the basic features of our system and our preliminary experiences towards its clinical implementation.

1 Introduction

In the context of cardiac interventions, minimizing invasiveness has inevitably led to more limited visual access to the target tissues. Minimally invasive approaches have emerged in response to the need to reduce morbidity associated with traditional techniques. Two key components of image-guided surgery (IGS) consist of first, the ability to register multi-modal pre- and intra-operative images to each other and to the patient, and secondly, to track instruments in real time during the procedure, while displaying them as part of a realistic model of the surgical volume. For the most part, IGS relies on the assumption that pre-operatively acquired images provide a sufficiently accurate representation of the intra-operative anatomy. This assumption is typically valid in areas such as orthopedic interventions, where the the anatomy primarily undergoes rigid transformations. However, it may not necessarily be valid for other soft tissue interventions, and it definitely poses a significant issue in the case of cardiac interventions. Considering the large myocardial deformations, intra-operative imaging using X-ray fluoroscopy, interventional MRI, US or video that provide accurate, real-time information on the anatomical changes are often employed in the modern operating rooms (OR) to ameliorate this situation. Moreover, stereoscopic, virtual and augmented reality techniques have been implemented to enhance visualization and surgical guidance.

We developed a novel surgical guidance package to facilitate therapy delivery on the beating heart. Our platform relies on trans-esophageal echocardiography

(TEE) for real-time intra-operative guidance, pre-operative cardiac models for anatomical enhancement, and magnetic tracking technology for real-time tracking of surgical instruments [1]. We envision that virtual reality (VR)-enhanced ultrasound (US) guidance will eliminate the need for intra-operative fluoroscopic imaging and allow the fusion of surgical planning and guidance [2, 3].

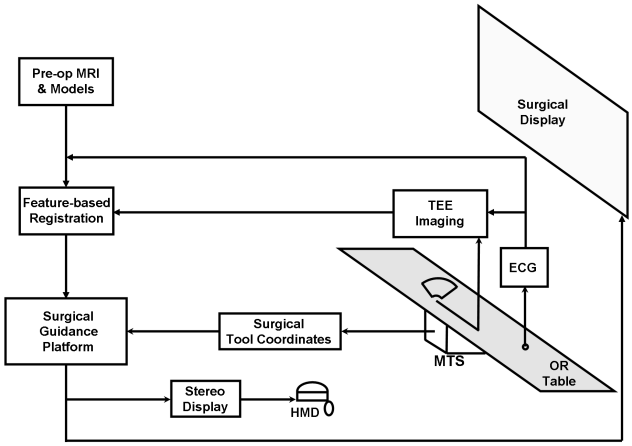


Fig. 1. Schematic representation of the OR setup during a VR-enhanced US guided procedure. Information is delivered to the surgeon either via overhead monitors or stereoscopically via head mounted displays (HMD).

To illustrate our long-term research endeavours, we follow a subject through the prospective clinical routine associated with this proposed technique. Pre-operatively, a dynamic cardiac MR dataset is acquired and used to generate a subject-specific heart model [4]. Peri-operatively, ECG-gated, dynamic real-time cardiac images are acquired via a tracked TEE probe, which are then post-processed into 3D volumes depicting the heart at different cardiac phases [5]. A feature-based registration technique is used to map the pre-operative model onto the the intra-operative anatomy revealed in US [6, 5]. Ultimately, the virtual surgical environment is augmented with models of the surgical instruments (i.e., the ablation tool) tracked in real-time using a magnetic tracking system (MTS) (**Fig. 1**). As a result, clinicians can explore the intracardiac environment in real-time via US, using the registered pre-operative models as guides, and navigate tracked surgical tools intuitively relative to the 3D cardiac anatomy.

This paper provides an overview of our augmented reality-based surgical guidance platform and focuses on the translation of the research from the IGS laboratory into the operating room. We discuss the clinical translation from the perspective of common intracardiac procedures, while outlining the necessary infrastructure and challenges that must be overcome to ensure smooth integration within the clinical environment in a modern OR.

2 Surgical Platform Architecture

Our surgical guidance platform integrates a wide variety components for IGS applications, including multi-modality image visualization, anatomical modeling, surgical tracking, and haptic control. Moreover, given the lack of visualization during off-pump surgery, we used this platform to build an AR environment to provide surgeons with a virtual display of the surgical field that resembles the real intracardiac environment to which they do not have direct visual access.

2.1 Augmented Reality Environment

Intra-operative Echocardiography: Ultrasound is widely employed as a standard interventional imaging modality. Specifically, 2D TEE provides good-quality images and eliminates the interference between probe manipulation and surgical work-flow, but the main drawback arises due to its inadequacy of depicting sufficiently crisp representations of the anatomical targets and surgical tools. While such inadequacies may be tolerable in *in vitro* studies performed in a laboratory environment under ideal conditions, they are highly amplified in a clinical setting by the complexity and variability of the anatomy, image quality, and orientation of the US beam with respect to the anatomy (**Fig. 2**). To address these limitations, we augment the 2D intra-procedure images with anatomical context supplied by the pre-operative models generated from CT or MR images acquired prior to the intervention.

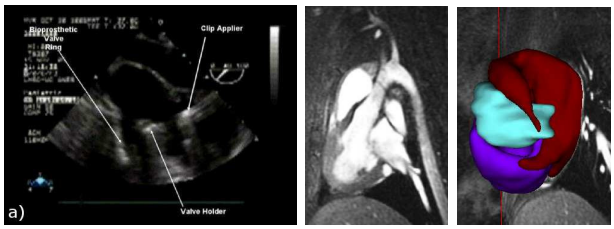


Fig. 2. a) 2D TEE image of the valve tool and clip applier inside a beating porcine heart; b) Pre-operative MR image and c) model of porcine heart at mid-diastole.

Pre-operative Anatomical Modeling: For *in vitro* studies involving cardiac phantoms, we rely primarily on pre-operative CT images for model building, given their good contrast and temporal resolution, and make use of automatic techniques for image segmentation, such as the Vascular Modeling Toolkit[4]. On the other hand, once translating to clinical porcine studies, the insufficient contrast between cardiac structures provided by CT without contrast enhancement

[4] VMTK: http://villacamozzi.marionegri.it/∼luca/vmtk/doku.php

forced us to resort to the use of MR images. Their excellent soft tissue characterization capabilities facilitated the of anatomical feature identification, leading to better quality subject-specific models. We first model each cardiac component, then assemble them together according to the complexity of the procedure. Typically, the main features of interest include the left ventricle myocardium (LV), the left atrium (LA), and the right atrium and ventricle (RA/RV).

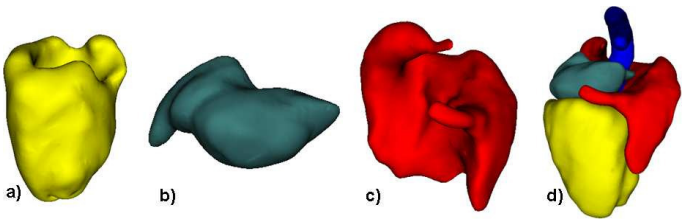


Fig. 3. Anatomical models of porcine heart chambers extracted from pre-operative MR images: a) LV; b) RA/RV; c) LA; d) Complete heart model: LV, LA, RA/RV.

However, MR imaging is still not the answer to all our doubts, as a next common obstacle is extracting fine anatomical structures (i.e. mitral or aortic valve) from low-resolution, thick-slice, clinical MR data. To ameliorate this inconvenience, we have developed and tested a technique on human data that relies on fitting a high-resolution, average heart model to clinical-quality MR datasets of new subjects via non-rigid registration, leading to reasonably accurate subject-specific models. Nevertheless, for the time being, porcine cardiac models are generated via manual segmentation, as we are in the process of building a porcine heart atlas similar to the human one.

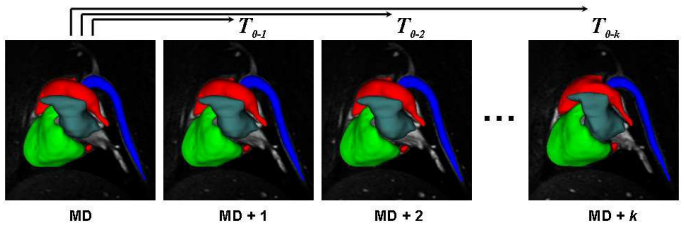


Fig. 4. Dynamic model of a porcine heart depicted at different cardiac phases.

Ultimately, a dynamic cardiac model depicting the heart at different time points in the cardiac cycle, is obtained by sequentially propagating the static model throughout the cardiac cycle with the motion estimated via non-rigid image registration [7] and rendered to portray the dynamics of the heart **Fig. 4**.

Pre-to Intra-op Registration: We employ a peri-operative feature-based registration technique to augment the intra-operative US images (**Fig. 5a**) with the pre-operative cardiac models (**Fig. 5b**). Easily identifiable targets in both datasets, the mitral (MVA) and aortic (AVA) valve annuli, are chosen to drive the registration. The pre-operative annuli are segmented manually from the cardiac MR image using a custom-developed spline-based segmentation tool (**Fig. 5b**) ; the intra-operative annuli are extracted from "pseudo" 3D US volumes generated by assembling 2D TEE images acquired at finite angular increments, according to their spatial and temporal time-stamps encoded by the MTS and ECG-gating. The algorithm first aligns the centroids (translational component) and normal unit vectors (rotational component) corresponding to the homologous annuli, and ultimately refines the alignment via an iterative closest point approach, minimizing the distance between homologous features.

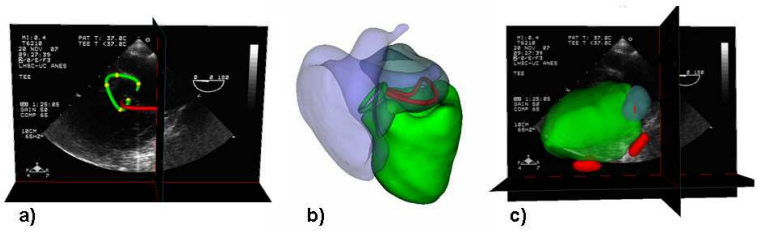


Fig. 5. a) Intra-operative US image, and b) pre-operative model, showing the MVA and AVA; c) Pre- and intra-operative datasets fused via feature-based registration.

This mapping technique is suitable for cardiac interventions, as the selected valvular structures are easily identifiable in both datasets, and they also ensure a good alignment of the pre- and intra-operative surgical targets. Furthermore, given the location of the features used to drive the registration, we expect good anatomical alignment in the surrounding regions, enabling us to employ this technique for a variety of image-guided interventions inside the heart.

Surgical Tracking: For all off-pump interventions, it is crucial for the surgeon to know the position and orientation of the instruments with respect to the intrinsic surrounding anatomy at all times during the procedure. In our platform, this feature is integrated via the NDI Aurora (Northern Digital Inc., Waterloo, Canada) MTS, which of course, imposes a series of limitations with respect to what equipment may or may not be present in the surgical field. Virtual representations of the instruments, registered to their physical counterparts, are tracked in real time relative to the pre- and intra-operative anatomy within the AR environment. In addition, a reference sensor is attached to a stationary region of the subject to avoid the need to recalibrate the "world" coordinate system in case of accidental motion of the subject or field generator. As an example, for a MV implantation, we need virtual representations for the TEE transducer,

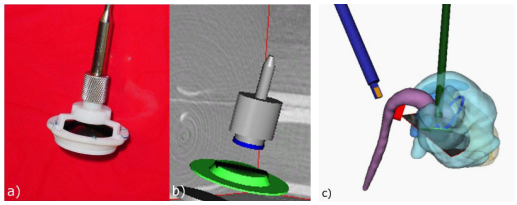


Fig. 6. a) Physical and b) virtual representation of a prosthetic mitral valve attached to the valve-guiding tool; c) AR environment specific to an *in vivo* porcine study.

valve-guiding tool and valve-fastening tool **Fig. 6**. A more sophisticated model is required for the US probe, as it incorporates the video feed from the scanner and the image plane must be automatically adjusted to changes in rotation angle and depth as manipulated by the sonographer.

3 Image-Guided Therapy Applications

Mitral Valve Implantation: Direct access was achieved using the Universal Cardiac Introducer (UCI) [8]. The AR environment consisted of the pre-operative model registered to the intra-procedure US, and virtual representations of the valve-guiding tool and valve-fastening tool — a laparoscopic clip applier. The procedure involves the positioning and fastening of the valve onto the native mitral annulus. Both steps entail the navigation of the tools to the target under guidance provided by the virtual models, followed by the correct positioning and attachment of the valve via surgical clips, guided via US (**Fig. 7**). Intra-operative assessment using Doppler imaging confirmed a successful valve placement, also observed in the post-procedure analysis. Nevertheless, one of the applied surgical clips failed to properly secure the valve due to poor penetration into the tissue.

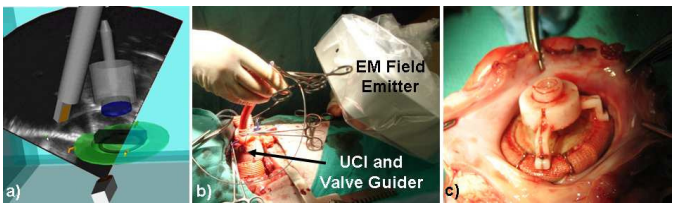


Fig. 7. a) AR environment showing virtual models of the US probe and surgical tools; b) OR setup during AR-guided interventions; c) Post-procedure assessment image.

Atrial Septal Defect Repair: The ASD repair procedure entails similar tasks as the MV implantation, however the surgical target is not readily defined. The

septal defect was created by removing a circular disc of tissue from the fossa ovale under US guidance, using a "hole-punching" tool. The blood flow through the ASD was identifiable on Doppler US. The repair patch was guided to the target under virtual reality guidance. Once on target, US imaging was used to correctly position the patch onto the created ASD and anchor it to the underlying tissue. Correct ASD repair was confirmed in the post-procedure assessment (**Fig. 8b**). While the VR-enhanced US guidance provided significant assistance in both navigation and positioning, this study also raised the need to improve tool design to better suit the limited intracardiac space.

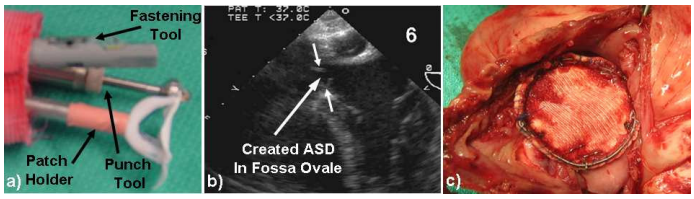


Fig. 8. a) Tools employed during the ASD creation and repair; b) 2D US image showing the septal defect; c) Post-operative image showing the successful ASD repair.

4 Roadblocks in Clinical Translation

In spite of our recent successes, the translation from the lab to the clinic has helped us identify caveats specific to the OR environment that had not posed major concerns within the *in vitro* laboratory conditions. Here we summarize some of these the lessons, hoping that some of you will learn from our experience.

How Accurate is Accurate Enough? One of the most frequent questions in IGS is "How accurate is your surgical platform"? One must keep in mind that, unlike orthopedic or neurological applications, cardiac IGS will always be prone to higher inaccuracies, mainly due to the high degree of rapid movement of cardiac tissue and the limitations of working in a blood filled environment, and amplified by the difficulties in building perfect anatomical models of the heart and the limited accuracy when mapping the pre-operative models to the intra-operative patient anatomy. We previously generated sufficiently accurate subject-specific models of the human heart capable to predict the MVA location throughout the cardiac cycle [4], and registered the models to their corresponding US images. Moreover, we quantified *in vitro* the alignment of the virtual models with their US images [9, 10], and assessed the accuracy in targeting static and dynamic "surgical targets" located on the epicardial surface of a beating heart phantom [11]. Given these applications, our system proves clinically feasible, considering the accuracy of the MTS on the order of 1-2 mm and the scope of the pre-operative models to enhance intra-operative anatomical context.

Zoom in where Needed: Surgical Target Identification. We currently employ the pre- and intra-operatively defined AVA and MVA to drive the model-to-subject registration peri-operatively in the OR. While the pre-operative features are better defined given the improved image quality, the intra-operative feature identification poses many challenges. We are currently exploring ways to acquire and reconstruct 3D US images from tracked 2D US images in nearly real-time, and employ automatic or semi-automatic algorithms to extract features. This approach would allow us to quickly and accurately identify the intra-operative surgical targets, and provide the flexibility to update the pre- to intra-operative registration as often as needed during the procedure. In spite of the recent launch of 3D TEE, this modality has a restrictive field of view which very often makes it impossible to visualize both the surgical targets and tools in the same volume; reconstructed 2D US, with its much larger field of view, is more robust if it can be achieved in near real time.

Cutting-edge Hardware: Surgical Tool Design. Surgical tool design and manufacture is a project of its own in IGS, especially when subject to multiple constraints imposed by the cardiac anatomy, interventional application, and surgical environment. We have always faced the challenge to design and build our own tools or adapt other available tools for the application at hand. For example, we had to build various prototypes for the valve- or ASD patch-guiding tools, embed the MTS sensors inside the tools, and test their compatibility with the surgical environment. Moreover, we adapted a laparoscopic stapler for abdominal interventions as a fastening device for both the valve and ASD patch, in spite of its slightly over sized dimensions for intracardiac applications. Most off-the-shelf tools do not comply with our requirements, and we do believe that the most efficient approach would be to involve a medical device manufacturer into the project and have them build tools that fit the applications. After all, sub-millimeter accuracy is meaningless unless appropriate surgical instruments are available for safe therapy delivery.

Minimizing our "Footprint" in the Operating Room. A "busy" environment is not uncommon in an OR, raising the concern of using magnetic tracking technology for surgical tool tracking as opposed to an optical system, which in turn implies the avoidance of any ferro-magnetic objects in close proximity to the magnetic field emitter [12]. Moreover, given that tracking accuracy decreases away from the magnetic field emitter, it is imperative that the field generator be placed within a range of 20-30 cm from the most probable tool location . Initially we had the field generator overhanging just above the surgical field, but space is also "expensive" in the OR, especially when surgeons complain of not having sufficient room in doing what they are best at. As this setup was obstructing the regular "task flow", we adopted the approach of embedding the magnetic field generator within the mattress of the operating table, underneath the heart. This location has the additional benefit of placing tracked sensors closer to the field generator than large bodies of metal such as rib spreaders, making the MTS

substantially more robust. Cables from the sensors are placed under or along the OR table to be as unobtrusive as possible. Similarly, the overhead monitors must be placed such that the surgical team can access them with minimal effort.

Making the New Look like the Old: Teaching AR to Surgeons. The AR guidance platform is a novel surgical technique and in spite of our efforts to make it a common part of the current operating rooms, we are well aware that it may not make its way into conventional surgery for a few years. Over the years, surgeons have become familiar with "standard views" of human anatomy. Although we are now able to provide an unlimited range of oblique views of tools and surgical targets, it is often best to stick to the displays surgeons are most familiar with to maintain intuitive guidance. The best approach is to make the new look as much like the old, and instead of overwhelming them with a lot of new technology at once, rather give them the time to get accustomed to the new environments, and let them tell you how impressed they are. After all, very few surgeons will give your new guidance platform the time of the day while still in development, but they will all want to be the first ones to perform the clinical trials, especially if highly ranked by their colleagues and widely advertised by a large medical device manufacturer!

Maximizing Returns: Optimally Display the AR Environment. Minimal invasiveness restricts both visual and surgical access. For off-pump intracardiac interventions, surgeons can't "see" what they do, and therefore the AR guidance platform **is** their eyes. We have considered several approaches regarding the most appropriate display of the AR environment to the surgeons, including a simple computer monitor, a flat screen overlaid onto the patient and located above the operating field, a stereoscopic screen enabling 3D visualization, or head-mount display (HMD) units, which allow the surgeons to directly "navigate" within a virtual volume [13, 14]. These different alternatives will be explored in our future work; for the time being, our surgical team has reported great comfort using both overhead monitors and HMD units for visualization.

5 Conclusions

We briefly described several IGS components developed and validated in our laboratory and outlined the path to transpose traditional cardiac surgery into an AR environment in a clinical setting. Some of the lessons learned while paving the road to the clinic have proved to be rather harsh and we still expect several other surprises down the road, given the increased complexity of the procedures attempted. But nevertheless, new image guidance tools are being developed almost every month, new generations of clinicians complete their medical training every year, so it really is not unrealistic at all to expect that operating rooms of the future will be readily equipped with image-guidance systems.

Acknowledgments. The authors thank Dr. Marcin Wierzbicki, Dr. Usaf Aladl and Dr. David Gobbi for assistance with software development and Sheri VanLingen and Louis Estey for technical support. In addition, we acknowledge funding for this work provided by NSERC, CIHR, ORDCF, OIT and CFI.

References

1. Linte, C.A., Moore, J., Wiles, A.D. *et al.*: Virtual reality-enhanced ultrasound guidance: A novel technique for intracardiac interventions. Comput Aided Surg. **13** (2008) 82–94
2. Wilson, K., Guiraudon, G., Jones, D. *et al.*: 4D shape registration for dynamic electrophysiological cardiac mapping. Volume 4191 of Lect Notes Comput Sci. (2006) 520–526
3. Wilson, K., Jones, D.L., Linte, C.A. *et al.*: Dynamic cardiac mapping on patient-specific models. In: Lect Notes Comput Sci. Volume 5241. (2008) 967–74
4. Linte, C.A., Wierzbicki, M., Moore, J. *et al.*: Towards subject-specific models of the dynamic heart for mitral valve surgery. Volume 4792 of Lect Notes Comput Sci. (2007) 94–101
5. Linte, C.A., Wierzbicki, M., Moore, J. *et al.*: From pre-operative cardiac modeling to intra-operative virtual environments for surgical guidance: An in vivo study. Volume 6918 of Proc. of SPIE. (2008) 69180D–12
6. Linte, C.A., Wierzbicki, M., Moore, J. *et al.*: On enhancing planning and navigation of beating-heart mitral valve surgery using pre-operative cardiac models. In: Proc. of IEEE Eng Med Biol Soc. (2007) 475–78
7. Wierzbicki, M., Peters, T.M.: Determining epicardial surface motion using elastic registration: Towards virtual reality guidance of minimally-invasive cardiac interventions. Volume 2878 of Lect Notes Comput Sci. (2003) 722–29
8. Moore, J., Guiraudon, G.M., Jones, D.L. *et al.*: 2D ultrasound augmented by virtual tools for guidance of interventional procedures. Volume 125 of Stud Health Technol Inform. (2007) 322–27
9. Wiles, A.D., Guiraudon, G.M., Moore, J. *et al.*: Navigation accuracy for an intracardiac procedure using virtual reality-enhanced ultrasound. Volume 6509 of Proc. of SPIE. (2007) 65090W–10
10. Wiles, A.D., Linte, C.A., Moore, J. *et al.*: Object identification accuracy under ultrasound enhanced virtual reality for minimally invasive cardiac surgery. Volume 6918 of Proc. of SPIE. (2008) 69180E–12
11. Linte, C.A., Wiles, A., Moore, J. *et al.*: Virtual reality-enhanced ultrasound guidance for atrial ablation: In vitro epicardial study. In: Lect Notes Comput Sci. Volume 5242. (2008) 644–51
12. Nafis, C., Jensen, V., Beauregard, L. *et al.*: Method for estimating dynamic EM tracking accuracy of surgical navigation tools. Volume 6141 of Proc. of SPIE. (2006) 61410K–16
13. Vogt, S., Khamene, A., Niemann, H., Sauer, F.: An AR system with intuitive user interface for manipulation and visualization of 3D medical data. In: Proc. of Medicine Meets Virtual Reality. Volume 98 of Stud Health Technol Inform. (2004) 397–403
14. Birkfellner, W., Figl, M., Matula, C. *et al.*: Computer-enhanced stereoscopic vision in a head-mounted operating binocular. Phys Med Biol. **48** (2003) 49–57

Accuracy of Tracked 2D Ultrasound during Minimally Invasive Cardiac Surgery

Andrew D. Wiles[1,3], Cristian A. Linte[2,3], Chris Wedlake[3], John Moore[3], and Terry M. Peters[1,2,3]

[1] Dept. of Medical Biophysics, The University of Western Ontario
[2] Biomedical Engineering, The University of Western Ontario
[3] Imaging Research Laboratories, Robarts Research Institute
London, Ontario, Canada
{awiles,clinte,cwedlake,jmoore,tpeters}@imaging.robarts.ca

Abstract. A 2D ultrasound enhanced virtual reality surgical guidance system has been developed in our laboratory. The system was tested in both the laboratory and the clinic. Recently, we studied the accuracy of two the ultrasound calibration methods with five different ultrasound transducers using a spherical object as the test platform. In this paper, we extend that work to use the superior vena cava and right atrium of a beating heart phantom as the metrological test artefact. The right atrium was imaged using the tracked ultrasound and the expected cross-sectional outline was determined using the intersection of the ultrasound plane and the surface model in the surgical guidance system. The expected and observed outlines are compared. The results show that the ultrasound calibration methods were sufficiently accurate in the spatial domain, but that temporal calibration is required to ensure accuracy throughout a given procedure.

1 Introduction

Traditional intracardiac surgeries are highly invasive procedures. The patient is opened at the sternum, placed on a cardiopulmonary bypass (heart-lung) machine and their heart is arrested so that the therapy can be applied. Risks to the patient include long recovery times, adverse immune response and possible neurological damage. To address these risks, we have developed an ultrasound enhanced virtual reality system whereby the therapy is applied to the beating heart using either (i) a mini-thoracotomy via the Universal Cardiac Introducer® (UCI) [1, 2] or (ii) via a percutaneous approach. The Atamai Viewer (`www.atamai.com`), a software package based on the Visualization Toolkit (VTK, `www.kitware.org`), is the foundation of the virtual reality software enhanced with real-time 2D ultrasound. The ultrasound transducer and surgical tools are tracked using a magnetic tracking system (Aurora®, NDI, Waterloo, ON, Canada) which, when employed in conjunction with a virtual environment, provides the surgeon with an interactive environment without direct vision.

Here we extend our previous work [3–5] on the evaluation of the performance of such systems. The system was tested for (i) targeting a static point source [3],

(ii) the insertion of a mitral valve in an ex-vivo porcine heart [4], and (iii) the identification of objects of point source and larger sized objects [5]. Here, we extend the work from [5] which used an approach similar to [6] except that both trueness and precision were assessed for the point source testing. Also, the 3D reconstruction used in [6] was ignored in favour of studying the difference between the outlines of the cross-section of a spherical object (table-tennis ball) appearing in the 2D ultrasound image, and the predicted outline of the spherical object based on the *a priori* model. In this paper, the spherical object is replaced with the inner surface of the superior vena cava and upper portion of the right atrium of a beating heart phantom (Chamberlain Group, Great Barrington, MA, USA) under static conditions. A geometrical model of the right atrium was created from a CT scan and the actual chamber was imaged using the ultrasound enhanced virtual reality surgical guidance system. The expected outline, computed using the intersection of the ultrasound plane and the geometrical model, was compared to the observed outline in the ultrasound image. The experiment was performed using two different ultrasound calibration methods (Z-bar [7] and phantomless [8]), and five different ultrasound transducers. We report on the agreement between the observed images and the expected outlines and provide a discussion on the level of performance observed.

2 Methods

The goal of this experiment is to test how well the tracked ultrasound image agrees with a virtual model of a static object[4]. In order to provide clinical context to the experiment, we chose to use a beating heart phantom as the metrological test platform. Several fiducial markers, whose positions were determined by identifying them with a tracked 3.2mm spherical tool tip, were placed on the heart surface and the mounting base for registration purposes (see Fig. 1).

A static CT scan of the heart phantom was obtained and the CT scan was segmented using the Vascular Modelling Toolkit (VMTK[5]) and ITK Snap. Two surface meshes were segmented from the CT volume for (i) the epicardial surface and (ii) the superior vena cava along with the upper portion of the right atrium. The surface model of the superior vena cava and the right atrium is provided in Fig. 1.

The beating heart phantom was placed in a 7% glycerol solution bath so that the various ultrasound transducers operated at the appropriate speed of sound (1540m/s). The images and surface models were loaded into the surgical guidance system and the models were registered with the phantom heart by digitizing the fiducial markers using the magnetic tracking system.

A 6 DoF magnetic tracking sensor was placed on each ultrasound transducer. The transducer was spatially calibrated using first the Z-Bar [7] and then the phantomless [8] methods. Once calibrated, the ultrasound image was tracked and displayed in the surgical guidance system (see [1] for details). The cross-sectional

[4] Future work includes performing a dynamic test with a temporal calibration.
[5] VMTK: `http://villacamozzi.marionegri.it/~luca/vmtk/doku.php`

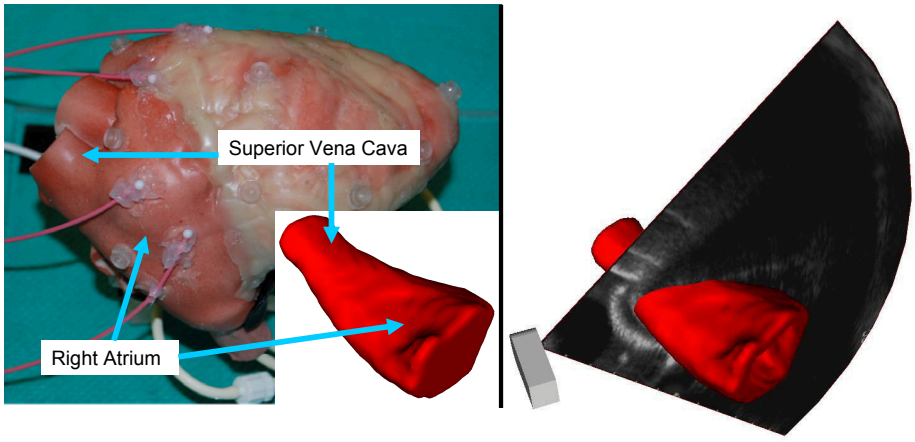


Fig. 1: Left, beating heart phantom and segmented surface of the superior vena cava and right atrium. Right, segmented surface being imaged by the Adult TEE where the expected outline occurs at the intersection of the ultrasound plane and segmented surface. Clear objects on the heart surface are fiducial markers. Pink flexcords with teflon balls on the tip are metrological devices for another study.

outline of superior vena cava and right atrium in the AP/LR plane were imaged at several points along the surface model. The expected outline was determined from the intersection of the infinite plane formed by the ultrasound fan and the surface model of the vena cava and right atrium using the `vtkCutter` class from VTK (see Fig. 1). The observed outline was marked on the ultrasound image itself. A sample ultrasound image using the adult transesophageal probe from Philips is provided in Fig. 2 where the expected and observed outlines are marked in green and yellow, respectively.

The points that define the outline form an irregular polygon. The centroid and area of the expected and observed polygons are computed by respectively averaging the center of mass and areas of a collection of triangles that were formed by the polygonal vertices. The expected and observed centroid positions and areas are compared and summarized in section 3.

3 Results

In all, two ultrasound calibration methods (Z-Bar [7] and phantomless [8]) were tested along with five different ultrasound transducers: (i) Aloka neuro echo transducer (N), (ii) Philips adult transesophageal echo transducer (AT), (iii) Philips pediatric transesophageal echo transducer (PT), (iv) Sequoia AcuNav intracardiac echo transducer (IC), (v) Aloka laparascopic echo transducer (L). In all, only 9 tests were performed because the Z-bar phantom used at our laboratory is too large for the laparascopic probe's field of view.

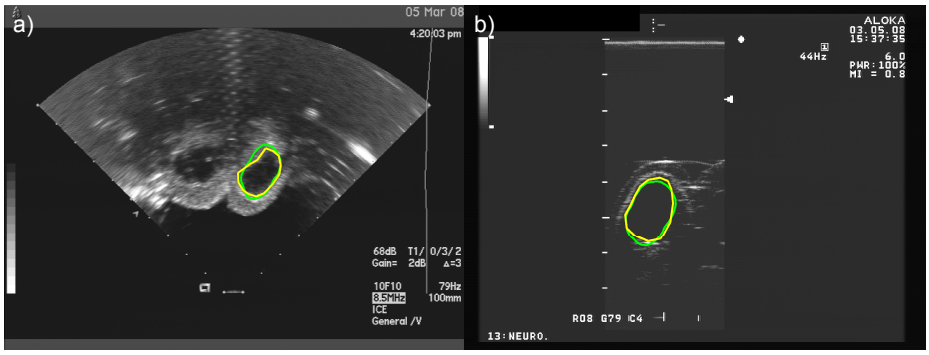


Fig. 2: Sample image of the right atrium with the expected (green) and observed (yellow) outlines for the a) intracardiac probe and b) laparascopic probe.

For each image of the right atrium surface model, the outline was marked with up to 16 points in the ultrasound image to form an irregular polygon. The centroid and areas of the expected and observed outlines were computed. The centroid distance error (d_{err}) is found by taking the magnitude of the difference between the observed outline centroid ($\boldsymbol{p}_o$) and the expected outline centroid ($\boldsymbol{p}_e$) given by $\|\boldsymbol{p}_o - \boldsymbol{p}_e\|$. The area percent difference was computed using the expected area as the reference such that $\%_{diff} = 100 \cdot (A_o - A_e) / A_e$. The centroid distance error RMS and the area percent difference mean and standard deviation are summarized in Table 1. Fig. 3 provides two plots, one for Z-bar calibrations and one for phantomless calibrations, which show the distribution of the centroid error on the image plane. Fig. 4 shows the distribution of the centroid distance error using a box plot format. The box plots are a better method to examine these than the statistics provided in Table 1 since the distributions are one-sided resulting in a non-Gaussian distribution.

Table 1: Summary results for the centroid error and area percent difference.

	Z-Bar Calibration		Phantomless Calibration	
Transducer	Centroid RMS	Area % Diff	Centroid RMS	Area % Diff
N	1.82	8.6 ± 12.3%	2.18	3.1 ± 12.1%
AT	5.68	16.1 ± 11.4%	2.64	-4.7 ± 11.5%
PT	3.12	-26.3 ± 10.3%	4.38	-12.6 ± 24.0%
IC	1.07	-17.14 ± 10.0%	2.78	-18.0 ± 8.7%
L	-	-	1.79	13.8 ± 7.4%

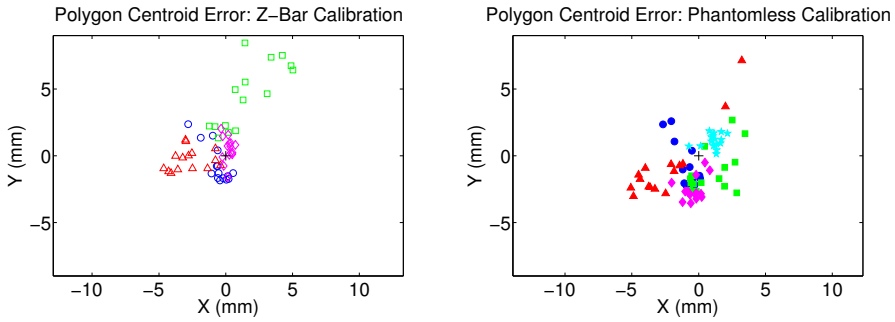


Fig. 3: Centroid Error. Centroid position error for observed and expected outlines are plotted on the image plane for the neuro transducer (○, ●), adult TEE (□, ■), pediatric TEE (△, ▲), ICE (◊,⧫) and laparascopic transducer (★). Hollow symbols: Z-bar. Filled symbols: Phantomless. Origin is marked with +.

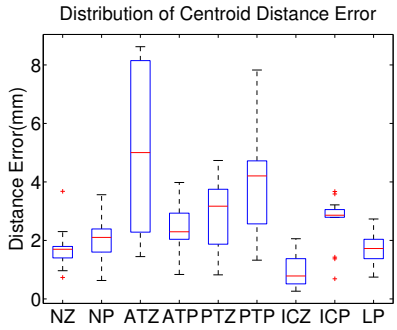


Fig. 4: Box plots showing the distribution of the centroid error for each test case.

4 Discussion

In our previous work [5], we noticed a significant difference in the transformations computed using the two different calibration methods. We attributed this to the need for the phantomless calibration tool to be held at multiple orientations relative to the ultrasound plane. For the experiments described here, we were able to set up the calibration such that the tracked calibration probe could be placed in a wider range of orientations. Therefore, the calibration matrices obtained from each calibration method were more similar, although the translations still differ by distances up to 3-4mm. Further investigation on how to optimize the two methods to ensure a proper calibration will be continued in the future.

In general, the results in Table 1 show that the two calibration methods perform well for our surgical guidance system. However, the centroid error RMS is quite high for the Adult TEE using the Z-Bar calibration as it approaches 6mm. The box plot in Fig. 4 shows the error ranging from less than 2mm to

more than 8mm. After examining some of the results we noticed a significant discrepancy in the data where some of the outlines matched poorly (Fig. 5a) and some them matched quite well (Fig. 5b). The Adult TEE with the Z-Bar was the first completed test case and some difficulty was experienced in holding the transducer steady, resulting in the ultrasound image moving during the data collection. This experiment was designed with the expectation that the data would be collected under pseudo-static conditions, i.e., the probe would be held steady at each collection. Hence, no temporal calibration was completed and the ultrasound video feed and the tracking system were not synchronized. Therefore, any movement of the transducer at the time of collection would result in an additional error such as the one seen in Fig. 5a. As the remaining test cases were completed, the user operating the ultrasound transducer became better at maintaining the pseudo-static requirements and the results improved. It is expected that our future work in this area will include a temporal calibration, particularly as we venture into experiments using a dynamic environment.

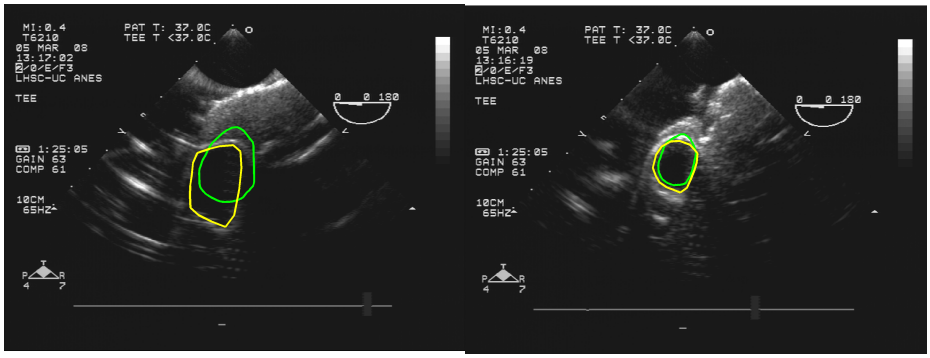


Fig. 5: Two examples for the Adult TEE probe using the Z-Bar calibration. In a) a translation between the expected and observed is seen but in b) the two are aligned quite well. The discrepancy is caused by the lack of a temporal registration and movement of the sensor in a).

We also noted some discrepancies in the surface model obtained from the CT and the actual cavity shown on the ultrasound image. Fig. 6a shows an image obtained where the expected outline is much smaller in area than observed. The enlarged view in the inset shows that there is no indication of any sort that the rear portion of the chamber exists in the ultrasound image. This is not due to the errors introduced by the lack of a temporal calibration, but is caused by the inability of the neuro ultrasound transducer to image the thin wall at the junction of the right atrium and pulmonary artery.

As shown in Fig. 6b, the outline observed using the pediatric probe was often smaller than expected. The point spread function in the transducer causes the boundaries of the right atrium to be blurred such that there is no clear indication as to where the boundary exists. The enlarged inset shows the level of blur

for this transducer and also explains why the area of the polygons was consistently smaller than expected. A similar effect was observed with the intracardiac transducer.

These discrepancies are a very good indication of why the ultrasound enhanced VR system is needed for accurate intracardiac surgery. In each of these cases the VR model provides an additional level of context for the actual shape of the vessel or chamber. This information would not be available in a procedure where only a 2D or 3D ultrasound image was used for guidance.

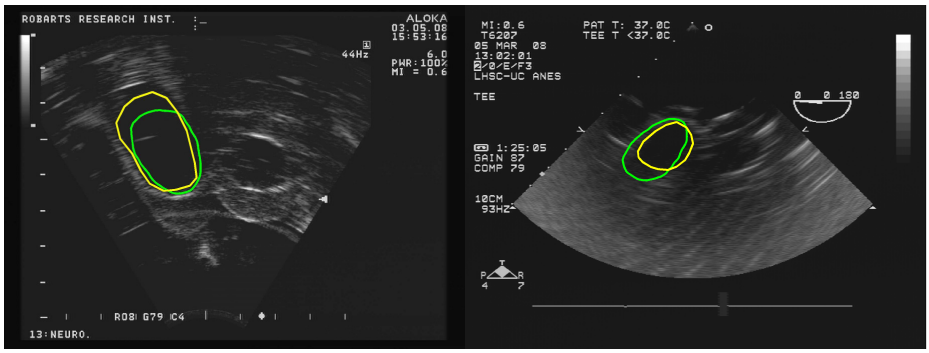


Fig. 6: Examples of errors caused by the ultrasound image for a) neuro probe and b) the pediatric TEE probe. The insets provide enlarged views of the right atrium without the expected or observed outlines. In a) the expected outline is much further away from the transducer than the observed outline. The image provided by the pediatric TEE in b) shows how blurred the surface edges become due to the point spread function of the ultrasound transducer.

In the results presented here the errors are based on a single ultrasound calibration for each test case. It would be beneficial if the uncertainty of the calibration could be modelled statistically whereby the calibration transformation matrix would have a 6×6 covariance matrix for the independent transformation parameters which would in turn provide an uncertainty model for the expected outlines. One method to determine the calibration uncertainty involves repeating the calibration 100 or 1000 times to generate a statistical sampling. However, this method is a daunting task where the time it takes to collect and mark the calibration images is significant. But, since both calibration methods fundamentally rely on a point-based registration, the uncertainty of the ultrasound coordinate frame can be estimated using target registration error (TRE) models [9,10]. The difficulty with this method is that determining accurate statistical models of the fiducial localizer error (FLE) is complicated. If we could estimate the FLE for each homologous point used in the registration, then the estimation of the calibration uncertainty would be trivial.

5 Conclusions and Future Work

It is clear that the two calibration methods are sufficient for calibrating the tracked ultrasound images, although it is critical that temporal calibration is also completed. Future work will include repeating this study under dynamic conditions and developing a statistical model that describes the uncertainty of a given ultrasound calibration.

Acknowledgements

The authors would like to thank Jacques Milner and Dr. Usaf Aladl at the Robarts Research Institute for their help in the image segmentation. This work was supported by CIHR, NSERC, ORDCF, CFI, UWO and NDI.

References

1. Linte, C.A., Moore, J., Wiles, A.D., Wedlake, C., Peters, T.M.: Virtual reality-enhanced ultrasound guidance: A novel technique for intracardiac interventions. Computer Aided Surgery **13**(2) (March 2008) 82–94
2. Guiraudon, G.M., Jones, D.L., Bainbridge, D., Peters, T.M.: Mitral valve implantation using off-pump closed beating intracardiac surgery: A feasibility study. Interact Cardiovasc Thorac Surg **6**(5) (Oct 2007) 603–607
3. Wiles, A.D., Guiraudon, G.M., Moore, J., *et al.*: Navigation accuracy for an intracardiac procedure using ultrasound enhanced virtual reality. In Cleary, K.R., Miga, M.I., eds.: Medical Imaging 2007: Visualization and Image-Guided Procedures. Volume 6509., SPIE (2007) 65090W
4. Linte, C., Wiles, A.D., Hill, N., *et al.*: An augmented reality environment for image-guidance of off-pump mitral valve implantation. In Cleary, K.R., Miga, M.I., eds.: Medical Imaging 2007: Visualization and Image-Guided Procedures. Volume 6509., SPIE (2007) 65090N
5. Wiles, A.D., Moore, J., Linte, C.A., Wedlake, C., Ahmad, A., Peters, T.M.: Object identification accuracy under ultrasound enhanced virtual reality for minimally invasive cardiac surgery. In Miga, M.I., Cleary, K.R., eds.: Medical Imaging 2008: Visualization, Image-guided Procedures. Volume 6918., SPIE (2008) 69180E
6. Rousseau, F., Hellier, P., Letteboer, M.M., Niessen, W.J., Barillot, C.: Quantitative evaluation of three calibration methods for 3-D freehand ultrasound. IEEE Trans Med Imaging **25**(11) (Nov. 2006) 1492–1501
7. Gobbi, D.G., Comeau, R.M., Peters, T.M.: Ultrasound probe tracking for real-time ultrasound/MRI overlay and visualization of brain shift. In: Medical Image Computing and Computer-Assisted Intervention - MICCAI 1999. Volume LNCS 1679. (1999) 920–927
8. Khamene, A., Sauer, F.: A novel phantom-less spatial and temporal ultrasound calibration method. MICCAI 2005, 8th International Conference Medical Image Computing and Computer Assisted Intervention **8**(Pt 2) (2005) 65–72
9. Fitzpatrick, J.M., West, J.B., Maurer, C.R.: Predicting error in rigid-body point-based registration. IEEE Trans Med Imaging **17**(5) (Oct 1998) 694–702
10. Wiles, A.D., Likholyot, A., Frantz, D.D., Peters, T.M.: A statistical model for point-based target registration error with anisotropic fiducial localizer error. IEEE Trans Med Imaging **27**(3) (March 2008) 378–390

Workflow Based Assessment of the Camera Augmented Mobile C-arm System

Joerg Traub[1], Seyed-Ahmad Ahmadi[1], Nicolas Padoy[1], Lejing Wang[1], Sandro Michael Heining[2], Ekkehard Euler[2], Pierre Jannin[3], and Nassir Navab[1]

[1] Chair for Computer Aided Medical Procedures (CAMP), TU Munich, Germany
[2] Trauma Surgery Department, Klinikum Innenstadt, LMU Munich, Germany
[3] INSERM / INRIA / IRISA, Faculte de Medecine, Universite de Rennes, France

Abstract. Several image guided surgery systems were introduced in the research community throughout the past decades. Only a few have found their way into the operating room and even a smaller number of them into everyday clinical routine. Within this paper, we describe a method for workflow based preclinical assessment of image guided surgery systems. This analysis can reveal surgical phase-related strengths and weaknesses of the system that can lead to design changes potentially leveraging future clinical acceptance. The method is based on a workflow analysis of the clinical procedure and the conduction of carefully designed surgery simulations resulting in a quantitative comparison of the current and the newly proposed system. The method was applied to assess the performance of the camera augmented mobile C-arm (CamC), which extends a standard mobile X-ray system by a video camera and provides an overlay of video and x-ray image without any online calibration or registration. This system was applied on several real-world and simulated vertebroplasty procedures and compared to conventional fluoro CT guided interventions. The analysis of these simulations provides initial quantitative results and a comparison of the current clinical method and the new system in terms of duration, X-ray exposure, and changes in the workflow.

1 Introduction

Radically novel systems for image guided interventions, e.g. systems using augmented reality visualization, face a long and tedious process from the conception of the initial idea until they are clinically applicable. This process includes the process of modeling the problem, designing the algorithms and the system, implementing and engineering the system, and finally evaluating the system in terms of its suitability for clinical usage.

One approach for the classification of the assessment of systems for image guided interventions was proposed by Jannin and Korb [1]. Their concept proposed an evaluation classified in six different levels. The here introduced model is motivated by this multilevel concept and proposes a reference-based method to compare computer assisted navigation systems with the currently used surgical procedure in the operating room. Observing, monitoring and analyzing the

medical workflow allow for the creation of surgery models. A simulated procedure is designed with respect to the surgical model and is used to compare the workflow of the novel approach against the currently applied surgical procedure. The comparison is based on a set of parameters monitored during both approaches. This will provide a structured way in assessing clinically relevant parameters already within a preclinical prototype and justify its further expensive and time-consuming ex-vivo and in-vivo studies. Thus already within an early stage of development an assessment based on this model shows potential to create meta models to assess parameters of the surgical strategy and performance by analyzing the workflow.

As an exemplary scenario for the proposed evaluation method, it was applied to the camera-augmented mobile C-arm system (CamC) which is further explained in section 2 and used for vertebroplasty procedures, further explained in section 4.1. For our exemplary evaluation, we monitor several parameters during the simulated vertebroplasties in order to verify two assumptions of the CamC system. One assumption is that it is reducing the radiation exposure of the surgical team and patients during the intervention. The other assumption is that it does not interfere with or complicate the current clinical workflow. Monitoring parameters relevant to these assumptions could prove the clinical usefulness of the system under inspection and justify the conduction of further cadaver and patient studies.

2 The Camera Augmented Mobile C-arm System

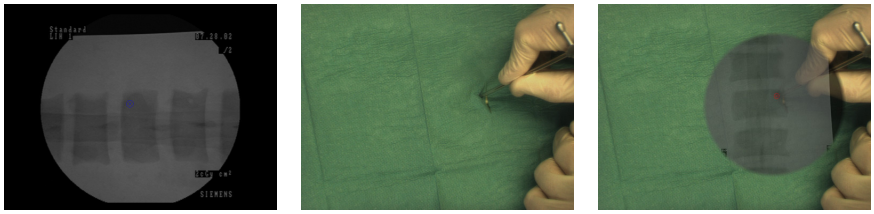


Fig. 1. The CamC system merges the the X-ray image (left) onto the video image (middle), creating an augmented view (right) which provides an intuitive interface for skin incision and instrument positioning.

Mobile C-arms are used every day in clinical routine in trauma and orthopedic surgery departments. With their introduction into the operating room, there was an increase in radiation exposure for both patient and surgical team [2]. Image guided surgery system and navigation solutions for trauma and orthopedic surgery aim at the reduction of the invasiveness including the radiation dose while ensuring the optimal patient outcome. Traditional navigation solutions for spine surgery use spatial localization systems to track instruments and patient in

3D space and register medical imaging data into the tracking coordinate frame [3, 4]. The used image guided surgery systems are thus either based on preoperative CT data (e.g. [5]) or intraoperative 3D (e.g. [6]) or 2D (e.g. [7]) all of them with their advantages and disadvantages. The systems based on intraoperative imaging acquire the image data within the tracking coordinate system and thus do not require any additional patient registration procedure. However, within clinical routine these systems require a long setup time, special instruments, add complex and manifold equipment to the operating room and they require technically trained surgeons or technical staff. The CamC system extends a standard mobile C-arm by a video camera (cf. figure 2) such that the video image and the X-ray image are overlaid (cf. figure 1) without further calibration and registration. This is achieved by a double mirror construction and a one time calibration routine during the construction of the device as described in Navab et al. [8]. The advantage of the system compared to traditional image guided surgery solutions is that it is entirely integrated in the C-arm and does not require any additional setup of hardware in prior to the intervention and no registration is required during the intervention. Since the system extends a mobile C-arm, it is furthermore possible to continue the surgery under standard fluoroscopic imaging anytime without delay. The system shows a promising extension to reduce the radiation exposure for the surgical team and patient.

3 Novel Method for Preclinical Comparison of Surgical Procedures

The method for structured preclinical evaluation of newly designed and developed image guided surgery systems is composed of five consecutive steps:

1. Workflow analysis of the clinical procedure. Within the operating room, several surgeries of the target application are recorded. For workflow analysis, already existing tools can be used (e.g. [9]).

2. Creation of a high level surgical model of the procedure. Recording and analyzing several surgeries, an average surgery and a surgical model with different hierarchical levels can be created. The system parameters in the top level should be representative of any performed procedure independent from the applied technology.

3. Creation of a simulated procedure and extraction of relevant parameters. Based on the clinical procedure, a phantom and a simulated procedure are designed. The major criterion is that the simulated procedure can be performed using any kind of image guided solution. The relevant parameters to be assessed during the simulated procedure need to be identified. Examples are the duration of each surgery phase, patterns in the workflow, accuracy and invasiveness of the procedure.

4. Recording and analysis of the simulated procedures. The designed assessment is performed on the phantom(s), comparing the current clinical procedure in the intervention room and the new system in the simulated setting. Using the same workflow analysis tools as for the clinical recordings, the quality

of the simulated procedure can be evaluated. This is done by comparing the clinical cases with the simulated procedures, providing an initial feedback on how close the simulation is to the real scenario.

5. Comparison of simulated procedures. The parameters and results of the simulated procedure under different image guidance modalities can then be compared directly. During the analysis, one has to consider the learning curve using a new system which may introduce a bias steming from untrained surgeons.

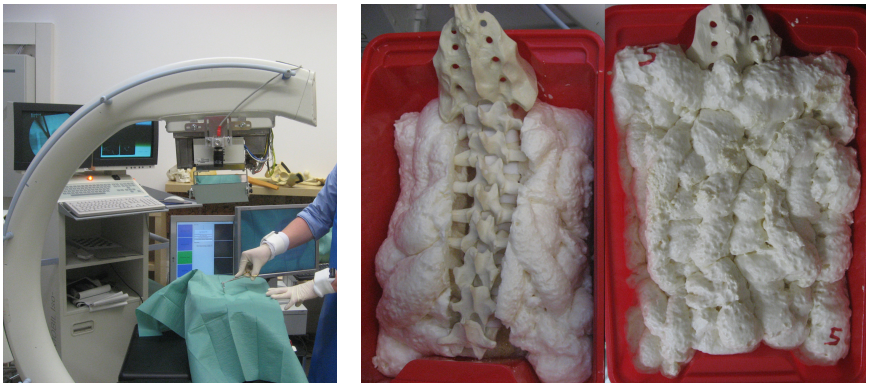


Fig. 2. The CamC system (left) and foam-coated phantom spines (right) used during the experiments.

4 Workflow Based Assessment of the CamC System

4.1 Workflow Analysis of the Clinical CT Fluoro Vertebroplasty Procedure

The clinical procedure for vertebroplasty is generally based on fluoro CT guidance (cf. figure 3). Within the CT intervention room of our medical partner institution, seven vertebroplasty procedures were recorded, each using two video cameras. One of the cameras was used to record the operation situs, the other one was used to record the interventionally acquired image data (fluoro and spiral CT). In our workflow analysis of seven vertebroplasty procedures performed by two different experienced surgeons, we observed an average duration of 31:14 [*min*]. The recorded videos were analyzed, synchronized and labeled with activity tags at a resolution of 1 Hz using a workflow analysis tool.

These procedures were further analyzed to derive the measurement criteria for a comparison of simulated vertebroplasty procedures using CT fluoro guided and the CamC system, as well as for the design of a simulated procedure that allows measuring the desired parameters. The major measurement criteria were the time required to complete specific tasks, the applied radiation dose during

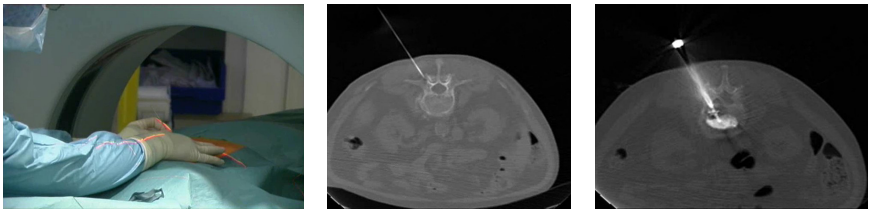


Fig. 3. Real vertebroplasty procedure using fluoro CT guidance (left). A K-wire or filling needle (middle) is positioned in the fluoro CT slice and controlled manually during insertion. Once the K-wire or filling needle is positioned inside the vertebra, the cement filling is performed using fluoro CT imaging (right).

specific phases of the surgery and the changes in the workflow and its complexity. Based on the workflow analysis, a surgical model for the vertebroplasty procedure was constructed. A high level description consists only of three phases (cf. figure 4). These phases and the fixed workflow events that indicate the phase borders are defined such that they can be uniquely identified, independent of the applied technology. These phases are:

1. System setup and target identification in the imaging data. This includes the setup of the equipment, devices, the positioning of the patient on the intervention table, and the localization of the identified vertebra within the image data and the identification of an access path within the imaging data. The skin incision indicates the end of this phase, independent from the imaging modality used.

2. Tool placement. This phase includes the placement of the instrumentation onto the inflicted vertebra and the process of inserting the instrument through the pedicle into the central vertebra cavity. It ends after a confirmation scan (e.g. spiral scan in CT) and as soon as mixing the cement compounds begins.

3. Cement augmentation. This includes the filling of the cavity of the vertebra with cement and frequent imaging updates, including a final confirmation scan, to monitor cement spreading in the vertebra. The overall intervention ends directly after the instrumentation is removed.

Several repetitive patterns could be observed in the recorded workflows. In particular, at several steps of the surgery, such as the positioning of the patient in phase 1 or the instrument insertion in phase 2, numerous fluoro scans have to be performed for immediate intra-operative validation of patient, vertebra and tool location. Such repeating patterns could be observed throughout all recorded real vertebroplasty procedures within the CT intervention room. They indicate potential optimization and improvements of the workflow through the introduction of computer assisted navigated procedures.

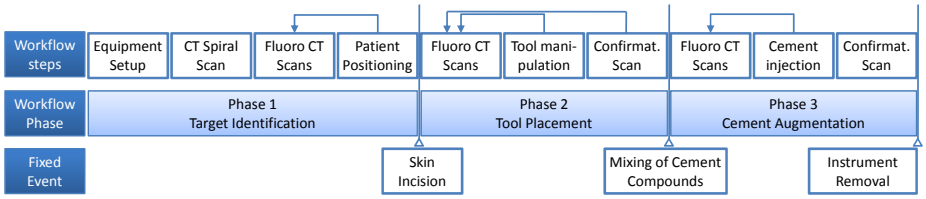


Fig. 4. Workflow analysis of vertebroplasty, consisting of a breakdown into 3 high-level phases. Several repetitive patterns (i.e. workflow loops) indicate potential workflow optimizations through computer-assisted navigation. The proposed system assessment method suggests to include such workflow patterns into the design of an evaluation procedure.

4.2 Design of a Simulated Surgical Procedure for Workflow Based Assessment in Vertebroplasty

An experiment and a phantom were designed to analyze the duration, radiation time and changes in the clinical workflow. We embedded five spine phantoms (T10-T12 and L1-L5) within a foam cover (cf. figure 2). For these phantoms, we simulated the complete process for vertebroplasty on the first lumbar vertebra (L1) as target anatomy using the CamC system and the twelfth thoracic vertebra (T12) target anatomy using fluoro CT navigated vertebroplasty. The anatomy of the L1 in the phantoms is only slightly different compared to the anatomy of T12, but is identical between all five phantoms. There was no variation in the anatomy of the same vertebra, however they were slightly differently embedded into the foam (orientation and location). Using the above defined transitions, we were able to assess the parameters in all phases independently. The last phase of cement augmentation was not representative during the experiments since it is not possible to perform realistic simulation in vertebrae without clinical indications and without cavity for the filling. The cement filling is also not subject for a navigated insertion, but controlled by real time imaging and image processing methods.

4.3 Workflow Analysis of Simulated Vertebroplasty Procedure using Fluoro CT and the CamC System

The simulated vertebroplasty procedure was performed five times for the fluoro CT guided method in the CT scanner room and five times using the CamC system (cf. figure 2). These procedures were recorded with cameras and analyzed to measure the above defined criteria. The overall duration, the individual phase duration in seconds as well as the overall radiation counted as number of scans were labelled from video. Finally, workflow patterns were derived and interpreted.

The results are shown in tables 4.3 and 4.3. The overall duration was longer using the CamC system. The average duration for the system setup and guided

insertion was 10:37 [*min*] for the CamC compared to 4:11 [*min*] using the CT fluoro procedure. This is mainly due to the non-routine usage of the newly developed system in contrast to frequent clinical usage of fluoro CT as well as multiple orbital rotations of the C-arm during the instrument insertion phase necessary for lateral X-ray images showing insertion depth. The radiation was reduced using the CamC system compared to the fluoro CT guided procedure. The exposure times of the C-arm are compared with the number of fluoro CT and spiral CT scans. Within these preliminary experiments, it was hard to quantify and compare the applied radiation dose between the CT scanner and the C-arm system. Therefore, we will integrate an external dose estimation device into the evaluation process for the upcoming cadaver experiments and clinical trials. This should result in an objective measure showing a considerable reduction of the radiation. Without the spiral CT scan for initial placement, the fluoro CT guided procedure would most probably have required a similar amount of X-rays.

The accuracy of the placement was measured by an observer not involved in the simulated surgeries and thus without a priori knowledge about the operative technique, based on post-interventional CT scans. The clinical classification was performed similar to the method of Arand et al. [10] for pedicle screw placements. Group A is classified by central screw positions without any perforation in any direction. Group B is classified by lateral, medial, caudal and cranial perforation smaller than the depth of thread of the screw. In our case, for canula insertion, we choose the perforation to be smaller than the needle thickness. Group C originally classify the screws with a perforation larger than the depth of thread of the screw. We used a perforation larger than the canula thickness as measurement criteria for the group C classification. In two of the five experiments using the CamC system, there was a medial perforation of the pedicle, three pedicles were placed perfectly. The medial perforation was caused by an undetected motion of the patient anatomy (phantom) that we could observe in the videos using the workflow analysis tool. An online detection of markers that are simultaneously visible in the video and the X-ray image was integrated after these experiments to detect any misalignment or motion during the procedure. Within the fluoro CT guided procedure, similar placement accuracy was observed. The fourth experiment showed medial perforation of the pedicle and in experiment five, the surgeon required a second attempt to place the needle within an acceptable position.

Observing workflow patterns, the new CamC system showed similar patterns as the simulated procedure with the fluoro CT technique. There was a similar amount of repeating patterns within the system setup phase and the instrument placement phase.

5 Discussion and Conclusion

We demonstrated a method for reference based assessment of a newly developed image guided surgery system and the clinically applied state-of-the-art proce-

	#1	#2	#3	#4	#5	Mean
duration without filling	10:36	15:10	13:20	6:15	7:44	10:37
setup time	2:31	0.38	0:20	0:45	2:33	1:22
insertion time	8:05	14:32	13:00	5:30	5:11	9:15
filling time	4:24	5:00	3:40	3:00	7:02	4:37
no. X-rays during setup	42	6	1	11	13	14.6
no. X-rays during insertion	13	37	38	16	19	24.6
placement category	A	A	A	B	B	

Table 1. Vertebroplasty experiment using the CamC system on five foam embedded spine phantoms (L1). Time in minutes:seconds, number of acquired X-ray images, and placement category according to Arand et al. [10].

	#1	#2	#3	#4	#5	Mean
duration without filling	3:30	5:25	4:45	3:05	4:09	4:10
setup time	0:54	1:00	0:30	0:35	0:30	0:41
insertion time	2:36	4:25	4:15	2:35	3:39	3:30
no. fluoro scans during setup	9	3	3	2	2	3.8
no. fluoro scans insertion	21	41	27	23	25	27.4
placement category	A	A	A	B	A (2nd try)	

Table 2. Vertebroplasty experiment using the fluoro CT based system on five foam embedded spine phantoms (T12). Time in mm:ss, number of acquired X-ray images, and placement category according to Arand et al. [10].

dure. We applied the proposed method in order to assess the CamC system for vertebroplasty procedures. The simulated procedure is a first step towards a reproducible reference-based assessment of the entire procedure. On the one hand, the simulation allows a comparison of the overall surgical procedure, here in terms of duration and applied radiation dose. On the other hand, it allows to assess the phases of a surgery independently. Most preclinical evaluations only consider the assessment of the navigated phase. This however can lead to a biased result, ignoring influences of the newly introduced system into other phases of the surgery (e.g. initialization and system setup). The here proposed method requires to assess the parameters independently from the used technology and on all levels of and aspects of the surgical procedure, as proposed by [1]. It should be mentioned that in this case, it was not entirely possible to compare the applied radiation dose, since no independent device measuring the radiation in the CT scanner and C-arm could be used yet. However, the simulation showed the potential of the CamC system in the reduction of the radiation dose. It also showed its current limitations resulting in an increased duration, mostly due to the surgeon's learning curve. Furthermore, the experiments show that there is no major modification within the workflow and its repeating patterns. Future work will focus on performing more objective and fairer simulations, by having a radiation measurement device and surgeons pre-trained on the new system. The presented method and experiments, however, provide an initial work towards

a standardized and objective evaluation of procedures for new image guided surgery solutions at a preclinical stage.

Acknowledgements : The authors would like to thanks Siemens Healthcare for partial financial support of the project.

References

1. Jannin, P., Korb, W.: Image-Guided Interventions - Technology and Applications. Volume chapter 18. Springer (2008)
2. Mehlman, C., DiPasquale, T.: Radiation exposure to the surgical team during fluoroscopy: "how far is far enough?". Orthop. Trauma **11** (1997) 392–398
3. Langlotz, F., Nolte, L.: Computer-assisted minimally invasive spine surgery - state of the art. In Mayer, H.M., ed.: Minimally Invasive Spine Surgery - A Surgical Manual. Springer (2006) 26–32
4. Holly, L.T.: Image-guided spinal surgery. The International Journal of Medical Robotics and Computer Assisted Surgery **2** (2006) 7–15
5. Jaramaz, B., A. M. DiGioia, I.: Ct-based navigation systems. In Stiehl, J.B., Konermann, W.H., Haaker, R.G.A., eds.: Navigation and Robotics in Total Joint and Spine Surgery. Springer (2003) 10–16
6. Euler, E., Heining, S., Riquarts, C., Mutschler, W.: C-arm-based three-dimensional navigation: a preliminary feasibility study. Computer Aided Surgery **8**(1) (2003) 35–41
7. Foley, K., Simon, D., Rampersaud, Y.: Virtual fluoroscopy: computer-assisted fluoroscopic navigation. Spine **26**(4) (2001) 347–351
8. Navab, N., Mitschke, M., Bani-Hashemi, A.: Merging visible and invisible: Two camera-augmented mobile C-arm (CAMC) applications. In: Proc. IEEE and ACM Int'l Workshop on Augmented Reality, San Francisco, CA, USA (1999) 134–141
9. Ahmadi, S.A., Sielhorst, T., Stauder, R., Horn, M., Feussner, H., Navab, N.: Recovery of surgical workflow without explicit models. In: Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention (MICCAI). (2006) 420–428
10. Arand, M., Hartwig, E., Hebold, D., Kinzl, L., Gebhard, F.: Präzisionsanalyse navigationsgestützt implantierter thorakaler und lumbaler pedikelschrauben. Unfallchirurg **104**(11) (2001) 1076–1081

The Visible Korean Human Phantom: Realistic Test & Development Environments for Medical Augmented Reality

Christoph Bichlmeier[1]*, Ben Ockert[2]**, Oliver Kutter[1]***, Mohammad Rustaee[1], Sandro Michael Heining[2]†, and Nassir Navab[1]‡

[1]Computer Aided Medical Procedures & Augmented Reality (CAMP), Technische Universität München, Germany
[2]Trauma Surgery Department, Klinikum Innenstadt, LMU, München, Germany

Abstract. This paper reports on the preparation, creation and first applications of the Visible Korean Human Phantom - VKHP that provides a realistic environment for the development and evaluation of medical augmented reality technology. We consider realistic development and evaluation environments as an essential premise for the progressive investigation of high quality visualization and intra operative navigation systems in medical AR. This helps us to avoid targeting wrong objectives in an early stage, to detect real problems of the final user and environment and to determine the potentials of AR technology. The true-scale VKHP was printed with the rapid prototyping technique "laser sinter" from the Visible Korean Human CT data set. This allows us to augment the VKHP with real medical imaging data such as MRI and CT. Thanks to the VKHP, advanced AR visualization techniques have been developed to augment real CT data on the phantom. In addition, we used the phantom within the scope of a feasibility study investigating the integration of an AR system into the operating room.

Key words: Medical Augmented Reality, Focus & Context AR Visualization, Test Beds for Augmented Reality

1 Introduction

In this paper, we propose a new genre of phantoms. The Visible Korean Human Phantom - VKHP provides a realistic environment to develop and evaluate future visualization techniques and surgical navigation systems taking advantage of augmented reality technology.

Beside the close collaboration among physicians, engineers and computer scientists within an interdisciplinary lab space, which we call *Real World Lab*, to

* bichlmei@cs.tum.edu
** ben.ockert@med.uni-muenchen.de
*** kutter@cs.tum.edu
† sandro-michael.heining@med.uni-muenchen.de
‡ navab@cs.tum.edu

investigate the needs of the final user and the potentials of new technology, we determined the importance of a realistic phantom for developing and evaluating our approaches as an important factor for progressive research. Experimental arrangements such as cadaver, animal or in-vivo studies to develop, showcase and evaluate AR technology are expensive, difficult to be justified and hard to obtain. Custom-made phantoms are often presented by the community to simulate a minor anatomic part or conditions. However, this can either produce artificial problems that are not present on real conditions or important issues are disguised. Both lead consequently to a non optimal system design and configuration.

It is extremely difficult, maybe even impossible, to build a phantom that corresponds to the physical properties of the human anatomy. Different companies like Gaumard, Laerdal or Meti offer manikins simulating different training environments for teaching medical emergency teams and particular medical situations like birth simulation. However, there is no solution yet on the market that provides a mannequin coming with real imaging data sets that is adapted to the needs for AR applications.

Many applications follow the strategy of using CT or MRI imaging data of unrealistic, custom-made phantoms to design a navigation or visualization system. Quite often, the produced phantoms and corresponding imaging data are capable of evaluating a certain task such as accuracy and time duration of instrument guidance. However, such imaging data excludes information that can be used in real conditions or includes information that misleads the user and results into distorted results that can not be projected onto real medical cases.

For instance, Sauer et al. [14] use a cantaloupe and a box filled up with mashed potatoes for the evaluation of AR guided needle biopsy. Birkfellner et al. [4] present a phantom skull with jelly that "was covered with polypropylene dust in order to make the jelly opaque" to simulate navigated neurosurgery. Traub et al. [16] built a wooden box for navigated surgical drilling that is equipped with metal spots serving as target spots and filled up with silicone. Sielhorst et al. [15] took a CT scan from a thorax phantom with nothing except a plastic spinal column inside the body to evaluate the quality of depth perception of different visualizations modes. We [1] built a phantom for pedicle screw implantation in spine surgery consisting of replaceable vertebrae. The CT scan of vertebrae surrounded by silicone and peas was then used for the evaluation of an AR navigation system.

Cadaver and animal studies come close to the real conditions in the OR. However, only few researchers having access to interdisciplinary facilities report on such lab and evaluation environments [17, 6, 3]. For the major part of the community, such experiments are not accessible at all. Only few groups report on evaluations on real patients [10, 4, 7, 5]. Unfortunately, this most realistic test bed for medical AR applications is far from being frequently accessible for continuous development.

2 Method

A major research topic in medical augmented reality addresses the improvement of visualizing medical imaging data superimposed on the patient. Naive superimposition of virtual anatomy on real objects such as the patient skin would result in misleading depth perception since the visualized data then appears to be outside rather than inside the body. We investigated a promising approach to provide the user of a video-see through head mounted display (HMD) [3] with an intuitive view through the transparent skin into the patient's body. For this reason the transparency of the skin region within the video images are manipulated due to geometric properties of the skin taken from the CT scan, the observer's position and view direction. They carried out a phantom, a cadaver and an in-vivo study and presented the results at [3]. Regarding the expensive preparation and execution of these studies shows that there is a major need for new test bed solutions.

A new approach for real-time visualization [11] and new graphics hardware allows us to augment a similar AR scene with direct volume rendering of CT data. Regarding consecutive and progressive development and evaluation of this high quality rendering approach, an earlier used thorax phantom [15, 3] can not measure up to the potentials of new rendering features anymore. Cadaver as well as volunteers can still serve for feasibility studies, however, both can not be provided for a permanent testing environment.

2.1 Visible Korean Human Phantom

Beside the *Visible Human Project* (Caucasian) (VHP) and the *Chinese Visible Human* (CVH) [18], the *Visible Korean Human* (VKH) [13] project provides full body imaging data consisting of MRI, CT and a photographic data set. We created a phantom from the CT data set of the VKH and prepared a set of operation sites for frequent applications in trauma surgery such as hip, shoulder, spine and neurosurgery. Such operation sites can be used for training surgeons in standard procedures within an almost realistic setting and developing new solutions for AR based image guided surgery in different surgical disciplines.

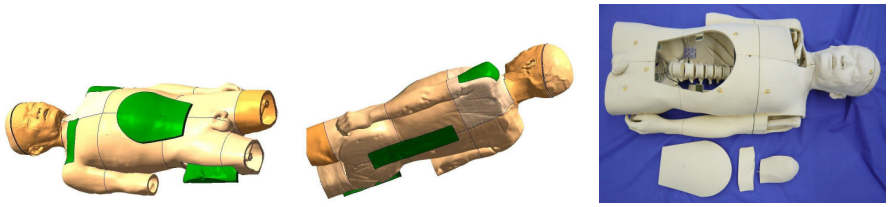
Five removable windows were fabricated into the phantom's skin (see Fig. 1(a)). This includes a 30x10 cm approach at the back of the phantom for the simulation of dorsal instrumented spine operations, for instance vertebroplasty, kyphoplasty or pedicular screw implantation and fixation. In addition, windows were integrated at the proximal shoulder and the proximal femur for internal osteosyntheses procedures, for instance plate and screw fixation. At the right clavicular region a 10x3 cm approach was built in to allow intramedullary clavicular fixation as well as intrathoracal drain application. Bearing in mind the soft tissue preparation defect, all drill and fixation steps can be obtained and guided by a navigation system.

The VKHP consists of the skin and bone structure from the CT scan, which is suitable for providing a test bed for different procedures in navigated hip, shoulder, spine and neurosurgery. First, CT data of the VKH was prepared using

the software 3-Matic and Mimics of Materialise [1] to create surface models of the anatomy to be printed in 3D. The required surface model format is STL [2] to be applicable for rapid prototype 3D printers. The virtual phantom was divided into components of maximum size of 700mm x 380mm x 590mm for the printing machine. In general, there are machines on the market that provide bigger print volumes. However, the components were prepared to be removed and plugged together for being able to adjust the VKHP according to a particular application and also for sharing parts of it among different lab spaces.

Our true-scale *Visible Korean Human Phantom* (VKHP) (see Fig. 1(b)) ranging from the head to the hips was printed from the prepared STL data (see Fig. 1(a)) with additive rapid manufacturing technique, *selective laser sintering*, by the company FIT [3]. The VKHP has been printed in layers (150 μm steps). Depending on the quality of the STL surface data, a theoretic level of detail of 0.7mm can be achieved.

To enhance the realism of the skin, the phantom was coated with skin colored powder, which also avoids non realistic specular highlights on the raw material.



(a) VKH data prepared in STL format. A set of windows (green) to operation sites have been prepared.

(b) Visible Korean Human Phantom

Fig. 1. The Visible Korean Human Phantom (VKHP) prepared with different operation sites for future experiments related to different procedures in navigated hip, shoulder, spine and neurosurgery.

2.2 Registration

The used tracking system localizes objects like surgical instruments and the patient equipped with a tracking target of passive infrared retro-reflective marker sets [3]. For high quality registration of the virtual CT and MRI data with the VKHP, we attached Beekly[4] CT markers to the skin of the phantom. The CT markers were coated with an infrared reflective foil in order to make them visible for our infrared camera tracking system described in detail at [3]. Next, point

[1] Materialise Group, Leuven, Belgium
[2] Standard Triangulation Language
[3] FIT Fruth Innovative Technologien GmbH, Parsberg, Germany
[4] Beekly Inc.

correspondences between the CT spots segmented from the CT data volume and the same spots detected by the tracking cameras are determined in order to compute a rigid registration transformation [3].

For practical reasons, we decided to divide the VKH data set into two separate tracked anatomic regions, the thorax/hip region and the head region. This allows us to reduce work space in our AR test bed and share the VKHP for parallel work.

The original VKH data set was acquired without the mentioned CT spots or comparable artificial landmarks that can be used for registration. For this reason, we took a high resolution scan from the two separated anatomic regions of the VKHP mentioned above. After the scan, we extracted the centroids of the attached CT markers from the new data set fully automatically, based on intensity thresholding and moments analysis. Finally, the original VKH CT data and the new phantom CT data are registered by non-linear intensity based registration. Mutual information was selected as a similarity measure and a best neighbor optimizer was used [8].

3 Results

Recent investigation in applying adaptive focus and context visualization for Medical Augmented Reality have shown promising results [9, 12, 3]. We are working on an approach where the transparency of the skin region in the video image is manipulated by geometric properties of the CT scan in the skin region, the observer's view direction and his/her point of view [3].

The VKHP has been successfully used for developing an improved and more flexible approach [11] for in-situ visualization of the VKH CT data set. The resulting AR scene corresponds to almost real conditions. Figure 2 shows focus & context visualization of the bone structure in the head and the thorax region on the VKHP. Within the long term project *Augmented Reality Aided Verte-*

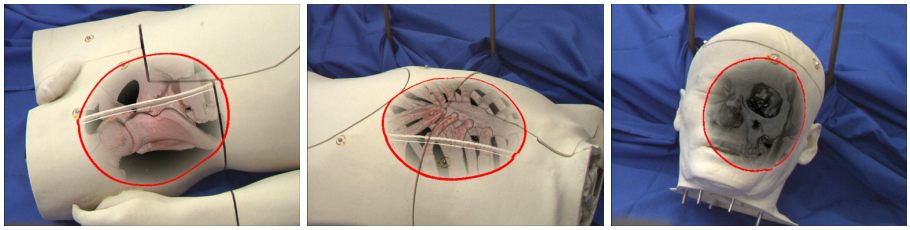
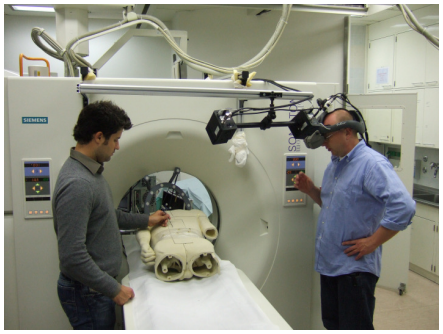


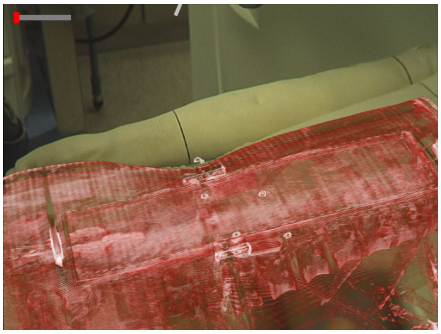
Fig. 2. First results of a new direct volume rendering approach for AR focus & context visualization augmenting CT data of the VKHP.

broplasty - ARAV [2], we use the VKHP for feasibility studies to integrate an AR system based on the stereo video see-through HMD into the trauma room of our clinical partner (see Fig. 3(a)). For the feasibility study (see Fig. 3(b)),

we do not augment the original data set but the CT from a test run. In this case we are interested more in planning the operating scheduling and equipment positioning than in visualization issues. Due to the true scale VKHP we are able to plan different camera setups, scan volumes and marker positioning with a dimensionally realistic experimental setup.



(a) VKHP is positioned on the bedding of the CT scanner.



(b) A trauma surgeon is observing the augmented VKHP.

Fig. 3. The Visible Korean Human Phantom (VKHP) used for a feasibility study with the Augmented Reality Aided Vertebroplasty - ARAV project.

4 Discussion

The interior of the VKHP is equipped with the bone structure printed from the VKH CT data set. This allows for simulating different procedures related to spine, hip and shoulder surgery. However, it does not provide a test bed for soft tissue applications such as tumor resection in different organic regions. There are rapid prototyping 3D printers on the market, which are capable of printing all kinds of materials with varying physical properties. The VKH data includes a set of segmented data volumes showing the main organs such as heart, colon, lung and liver. For a second version of the VKHP, we plan to install different soft tissue organs into the phantom that can also be removed when not needed or replaced when being damaged due to experiments.

Feedback from physicians of different disciplines such as anatomy, anesthesia and trauma surgery recommended using the AR system and the phantom simulation scenarios to train medical students and professionals. Different issues of teaching in anatomy courses can be addressed by the present AR system to study anatomic structures, inter-organic functionality, surgical access routes to operating sites and analysis of different imaging modalities.

5 Conclusion

We presented a new type of phantom adapted to the needs of the development and evaluation of visualization techniques and navigation systems in medical augmented reality applications. The Visible Korean Human Phantom - VKHP is printed from real CT data, which allows for accurate registration and realistic visualization of real patients data. This concept helps us to avoid cadaver and animals studies, which are expensive, often difficult to be justified and hard to be accessed. The VKHP has been used for the development of an advanced approach of direct volume rendering of CT data. In addition, we used the VKHP for a feasibility study to investigate the integration of our AR system into the trauma room. Following the statement of Park et al. [13] that "the VKH has exciting potential applications in the fields of virtual surgery, virtual endoscopy, and virtual cardiopulmonary resuscitation", we believe that the research progress in Medical Augmented Reality can be strongly stimulated by using comparable phantoms, since realism of the test and develop environments can be extremely increased. In order to support this evolution, we would like to share upon request the STL data, which is ready to be printed.

6 Acknowledgements

Special thanks to Frank Sauer, Ali Khamene, and Sebastian Vogt from Siemens Corporate Research (SCR) for the design, setup, and implementation of the in-situ visualization system RAMP they provided us. Thanks to Konrad Zürl and Oliver Wenisch from A.R.T. GmbH for providing cameras and software for the outside-in tracking system. We also want to express our gratitude to the radiologists and surgeons of Klinikum Innenstadt München for their precious contribution in obtaining CT data. Thanks also to Joerg Traub and the other members of the NARVIS group for their support.

References

1. C. Bichlmeier, S. M. Heining, M. Rustaee, and N. Navab. Virtually Extended Surgical Drilling Device: Virtual Mirror for Navigated Spine Surgery. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2007, 10th International Conference*, pages 434–441, Brisbane, Australia, October/November 2007.
2. C. Bichlmeier, H. Sandro Michael, R. Mohammad, and N. Nassir. Laparoscopic Virtual Mirror for Understanding Vessel Structure: Evaluation Study by Twelve Surgeons. In *Proceedings of the 6th International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 125–128, Nara, Japan, Nov. 2007.
3. C. Bichlmeier, F. Wimmer, S. M. Heining, and N. Navab. Contextual Anatomic Mimesis: Hybrid In-Situ Visualization Method for Improving Multi-Sensory Depth Perception in Medical Augmented Reality. In *Proceedings of the 6th International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 129–138, Nov. 2007.

4. W. Birkfellner, M. Figl, C. Matula, J. Hummel, H. I. R Hanel, F. Wanschitz, A. Wagner, F. Watzinger, and H. Bergmann. Computer-enhanced stereoscopic vision in a head-mounted operating binocular. *Physics in Medicine and Biology*, 48(3):N49–N57, 2003.
5. A. del Río, J. Fischer, M. Köbele, J. Hoffman, M. Tatagiba, W. Straßer, and D. Bartz. Intuitive Volume Classification in Medical Augmented Reality (AR). *GMS Current Topics in Computer- and Robot-Assisted Surgery*, 1, 2006.
6. M. Feuerstein, T. Mussack, S. M. Heining, and N. Navab. Intraoperative laparoscope augmentation for port placement and resection planning in minimally invasive liver resection. *IEEE Trans. Med. Imag.*, 27(3):355–369, March 2008.
7. W. E. L. Grimson, T. Lozano-Perez, W. M. Wells, III, G. J. Ettinger, S. J. White, and R. Kikinis. An automatic registration method for frameless stereotaxy, image guided surgery, and enhanced reality visualization. *IEEE Trans. Med. Imag.*, 15(2):129–140, 1996.
8. J. Hajnal, D. Hawkes, and D. Hill. *Medical Image Registration*. CRC Press, 2001.
9. D. Kalkofen, E. Mendez, and D. Schmalstieg. Interactive Focus and Context Visualization for Augmented Reality. In *Proceedings of the 6th International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 191–200, Nov. 2007.
10. A. P. King, P. J. Edwards, C. R. Maurer, Jr., D. A. de Cunha, D. J. Hawkes, D. L. G. Hill, R. P. Gaston, M. R. Fenlon, A. J. Strong, C. L. Chandler, A. Richards, and M. J. Gleeson. Design and evaluation of a system for microscope-assisted guided interventions. *IEEE Trans. Med. Imag.*, 19(11):1082–1093, 2000.
11. O. Kutter, A. Aichert, J. Traub, S. M. Heining, E. Euler, and N. Navab. Real-time Volume Rendering for High Quality Visualization in Augmented Reality. In *AMIARCS 2008*, New York, USA, Sept. 2008. MICCAI Society.
12. M. Lerotic, A. J. Chung, G. Mylonas, and G.-Z. Yang. pq -space based non-photorealistic rendering for augmented reality. In *Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention (MICCAI)*, volume 2, pages 102–109, 2007.
13. J. Park, M. Chung, S. Hwang, Y. Lee, D. Har, and H. Park. Visible korean human: Improved serially sectioned images of the entire body. *Medical Image Analysis*, 24(3):352–360, March 2005.
14. F. Sauer, A. Khamene, B. Bascle, S. Vogt, and G. J. Rubinob. Augmented reality visualization in imri operating room: System description and pre-clinical testing. In *Proceedings of SPIE, Medical Imaging*, volume 4681, pages 446–454, 2002.
15. T. Sielhorst, C. Bichlmeier, S. Heining, and N. Navab. Depth perception a major issue in medical ar: Evaluation study by twenty surgeons. In *Proceedings of MICCAI 2006*, LNCS, pages 364–372, Copenhagen, Denmark, Oct. 2006. MICCAI Society, Springer.
16. J. Traub, P. Stefan, S.-M. M. Heining, C. R. Tobias Sielhorst, E. Euler, and N. Navab. Hybrid navigation interface for orthopedic and trauma surgery. In *Proceedings of MICCAI 2006*, LNCS, pages 373–380, Copenhagen, Denmark, Oct. 2006. MICCAI Society, Springer.
17. F. K. Wacker, S. Vogt, A. Khamene, J. A. Jesberger, S. G. Nour, D. R. Elgort, F. Sauer, J. L. Duerk, and J. S. Lewin. An augmented reality system for mr image - guided needle biopsy: Initial results in a swine model. *Radiology*, 238(2):497–504, 2006.
18. S.-X. Zhang, P.-A. Heng, Z.-J. Liu, L.-W. Tan, M.-G. Qiu, Q.-Y. Li, R.-X. Liao, K. Li, G.-Y. Cui, Y.-L. Guo, and Y.-M. Xie. Chinese visible human data sets and their applications. In *HCI (12)*, pages 530–535, 2007.

Benefits of augmented reality function for laparoscopic and endoscopic surgical robot systems

Naoki Suzuki[1], Asaki Hattori[1], Makoto Hashizume[2]

[1] Institute for High Dimensional Medical Imaging, The Jikei University School of Medicine, 4-11-1 Izumihoncho, Komae-shi, Tokyo, Japan
{nsuzuki, hat}@jikei.ac.jp
[2] Center for the Integration of Advanced Medicine and Innovative Technology, Kyushu University Hospital, 3-1-1 Maidashi, Higashi-ku, Fukuoka, Japan
mhashi@dem.med.kyushu-u.ac.jp

Abstract. Various surgical robotics need augmented reality functions for the extension of their abilities. To enhance the benefit of augmented reality for robotic surgery, two research experiences are described. One is the loading of the augmented reality function onto the laparoscopic surgical robot. The results of the clinical case where the commercial surgical robot da Vinch loaded with augmented reality are described. The other is the endoscopic robot system which contained an augmented reality function. The results from an animal experiment where an EMR procedure on a pig was performed using augmented reality functions so that the location and direction of the robot's tip, which is free to move in the stomach, are detailed.

1 Introduction

1-1 Augmented reality for surgery works

There are two reasons why the amount of research on navigation surgery using augmented reality technique and its clinical applications have recently increased. First is that the capability of computers has increased to the extent that they can now display accurate real-time 3D images of the internal parts of patients. In the past, only the larger computers or computers that specialized in displaying images could accomplish this task. But now, even the most standard type of lap-top computers can show 3D images that hold information that will satisfy clinical needs. This has moved navigation surgery closer to reality. Secondly, although we still need optical markers or plural registration points, the technology for detecting the position of the target in real-time has improved and the price of commercial equipment has become cheaper. This has resulted in an increase in the number of people doing applied research on navigation surgery[1-6].

1-2 Robot surgery needs augmented realty

The development and applications of surgical robots has increased rapidly in the last 10 years[7-10]. The structure of the surgical robot also shows variations depending on the area of research. The surgical robot and augmented reality based on navigation surgery are deeply related to each other. That is, in case of robot surgery, the surgeon is always removed from the patient. For example in da Vinci, the most popular surgical robot for clinical cases, the surgeon operates the robot system from a surgeon's console positioned near the patient on the operating table. Most of the developed surgical robot systems have a similar structure where the surgeon has to control the surgical robot through a man-machine interface and can not see the operating field directly. Instead the vision is provided by a camera or some other optical system located near the operating field. Generally, the surgeon has to operate the surgical robot using this limited view compared with what is obtainable using the naked eye. Unfortunately, the detailed condition of the operation field and also the accurate direction of view are sometimes lost during this kind of operation. This happens frequently when an endoscopic surgical robot is used as it can bend its head to maneuver its robot arms and eyes (described in 2-2). That is, compared with other types of surgery such as open surgery or ordinary laparoscopic surgery, the robotic surgeon needs more information on the field of view. Safer and more rapid surgery can be achieved if the robotic surgeon has an indication of the targeted area or vessels near the operation field which were hidden from the real view. We describe the application of augmented reality technique for the laparoscopic robot and for the endoscopic robot, respectively, according to our research experience.

2 Method and results

2.1 Augmented reality for the laparoscopic robot (da Vinci)

This augmented reality system is a combination of a commercial laparoscopic robot (da Vinci) and a developed augmented reality system (Figure 1). The da Vinci's surgeon's console system has a stereo display and handles that manipulate the robot arm on the patient side. The developed augmented reality system for this robot consists of two devices. One is the optical location sensor (POLARIS, Northern Digital, Inc.) that measures the location of the laparoscope. The other is a graphic workstation (GWS, OCTANE MXE, Silicon Graphics Inc.) that captures the laparoscopic image from the da Vinci system and superimposes 3D organ models to that image. The superimposed image is outputted to the stereo viewer of the surgeon's console as a stereo paired image in real time. The optical marker is attached away from the laparoscope's tip so as not to interfere with the laparoscope's movement. Using the data of the location sensor, the GWS transforms the coordinate system of the 3D organ model to the surgical field coordinate system and renders the 3D organ model image. Fig.2 shows the arrangement of these systems in the operating room.

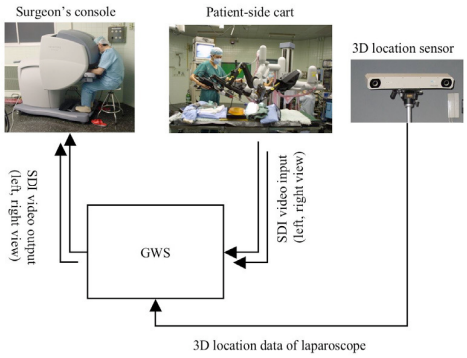


Fig. 1. System Configuration

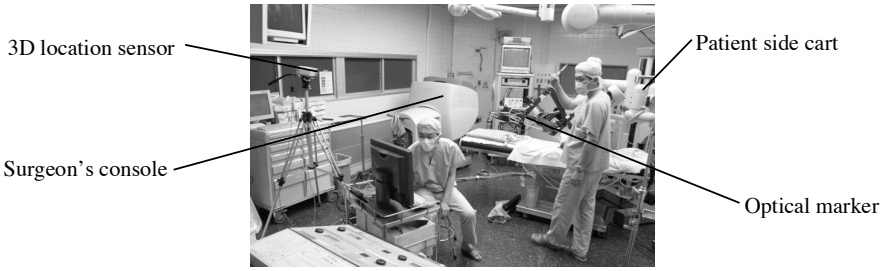


Fig. 2. Arrangement of the system in the operating room

For this system, we established following facilities. First, a display of organ models is updated in real-time following the movement of the laparoscope. Second, the surgeon is able to change the rendering style of the organ models according to the situation. Third, the system setup and calibration needs to be simple and quick to set up. Because the laparoscope needs to be sterilized, the location sensor marker can not be fixed to the laparoscope. Therefore, we attached the location sensor marker and calibrated the system just before the operation.

In order to fuse the laparoscope image and 3D organ models, the laparoscope and organ models coordinate system has to be transformed to a world coordinate system. After the laparoscope optical parameters were measured and registered the same locations on the organ model and the real organ, we calculated the transformation matrix for fusing the images. For the registration process, we measured four locations on the real organ by location sensor, and specified the same locations of the organ model on the GWS screen. Using the data from these locations, the transformation matrix was calculated.

After the registration, the 3D organ model image is superimposed onto the laparoscope's image. Figure 3 is the superimposed image of each (left and right) eye. Those 3D images are prepared for the cholecystectomy. The top window is the left eye's view and the bottom window is the right eye's view. Both are the view from under the liver. The gallbladder, common bile duct and hepatic duct model is

displayed. These superimposed images are outputted to the da Vinci surgeon's console.

After verification of the developed system through plural steps of an animal experiment, we examined the system during a cholecystectomy. Before this examination, we reconstructed the patient's organ models (liver, gallbladder, common bile duct and hepatic duct) from a 3DCT dataset. After attaching a sterilized optical marker to the laparoscope, we measured the laparoscope's parameters and registered the locations on the patient's organ model and real organ.

For the surgeon's usability, we used a foot switch to select whether to superimpose the organ model in the laparoscope's view or not. Also, the surgeon was able to change the organ model's transparency and color for distinguishing the organ model from the laparoscope's view.

Figure 4 shows a scene of the clinical experiment at a cholecystectomy. The superimposed image is shown in Figure 5. The gallbladder, common bile duct and hepatic duct model are superimposed onto the laparoscope's surgical field image. The frame rate at this experiment was 7-9 fps for each eye respectively. Figure 6a and 6b show the surgeon's console displaying the superimposed image as a stereo paired image.

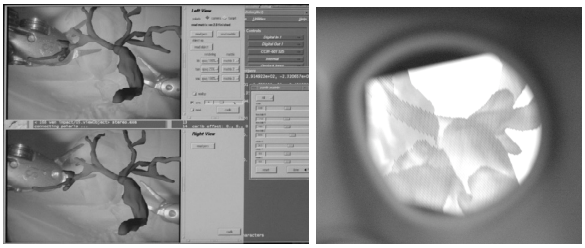


Fig. 3. Result images showing 3D organ models superimposed to the laparoscopic image; a: GWS display, top window: the left eye's view, bottom window: the right eye's view, b: a scene of the right eye's viewer of the surgeon's console

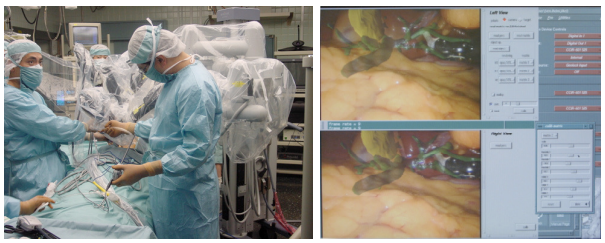


Fig. 4. A scene showing registration during a cholecystectomy (left). Superimposed images onto the GWS display (right). The patient's organ models (gallbladder, common bile duct and hepatic duct) are superimposed onto the surgical field image; top window: the left eye's view, bottom window: the right eye's view

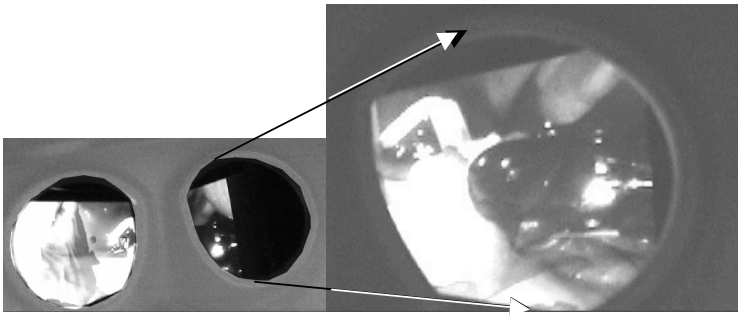


Fig. 5. Superimposed image on the surgeon's console. a: a scene of the stereo viewer from surgeon's console, b: a scene from the right eye's viewer.

2-2 Augmented reality for an endoscopic robot

This system is composed of a developed endoscopic robot system and an augmented reality system. Figure 6 shows the system outline. The endoscopic robot system has a structure whose distal part of the elastic tube has endoscopic eyes and a pair of small sized two robot arms on the both side of the eye. The distal part of each arm has forceps shaped endo-effecter. A wire drive mechanism is applied to operate the robot arms. The diameter of a robot arm is 3 mm and the maximum size of the distal part of the robot is 21 mm so that it is able to reach the gastric tube via a esophagus. The component power of three stainless wires enclosed inside a thin elastic tube controls the endo-effecter. An endoscopist performs the operation of the endoscope shaped body of the robot to move it into the gastric tube, and a surgeon beside the endoscopist remotely controls both robot arms with the controller panel.

The augmented reality system consisted of two visual devices. One is a graphic workstation (GWS: OCTANE MXE, Silicon Graphics Inc.) that has a digital video processing board installed in it. The GWS captures the endoscopic video image and superimposes a 3D organ model onto the captured video image. The surgeon is able to watch the superimposed video on the endoscope's monitor. One is the small magnetic location sensor (mini BIRD, Ascension Technology Co.) that always measures the location and the direction of the endoscopic robot's tip. The sensor is fixed to the tip of the robot (Figure 6). Using the updated positional data from the location sensor, the GWS transforms the coordinate system of the 3D organ model to the surgical field coordinate system and renders the 3D organ model onto the endoscopic captured video image. The 3D organ models are previously reconstructed from the patient MRI or CT dataset.

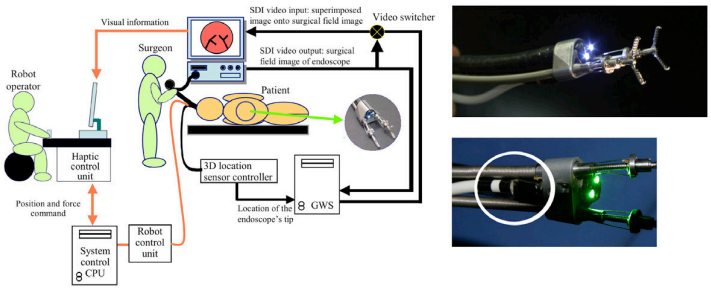


Fig. 6. System outline (left). The appearance of the endoscopic robot when attached to a 3D location sensor (right).

Fig. 7. A scene from an animal experiment (the left monitor displays a normal endoscope view, the right monitor displays a superimposed endoscope view that has a navigation function)

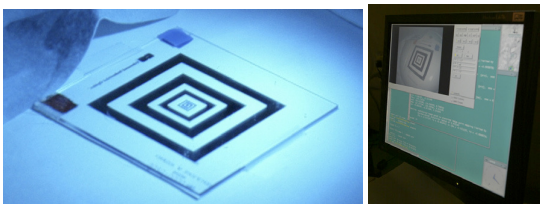


Fig. 8. Measurement of the endoscope's optical parameters (a: the reference board that is captured by the GWS, b: the GWS display, the captured reference board image is processed on the GWS)

Figure 7 shows a scene from the animal experiment (A pig, total weight was 30 kg) for the verification of the developed system. First, we measured the endoscope's optical parameters using the reference board (Figure 8). The captured reference board image was processed on the GWS, and the endoscope's parameters were expressed numerically. Second, we registered the coordinate system of the surgical field and the 3D organ model. The surgeon pointed at the real body surface locations using the

location sensor marker and specified the same locations on the 3D model on the GWS display. After registration, the surgeon inserted the endoscopic robot through the mouth cavity via the esophagus into the stomach. The 3D organ model that surrounded the stomach indicated the location and direction of the robot and was superimposed onto the endoscopic image (Figure 9). In this subjective image, the spine, ribs, liver and hepatic artery models were displayed. In this figure, the top left small window also shows an objective location of the robot's tip in the coordinate system of the organ model. The viewpoint of this window could be set interactively. The surgeon was able to change the transparency and the color of the organ models and switch from superimposed video to non-superimposed video, depending on the situation of the surgical field. The frame rate for this experiment was 12-14 fps. In this verified experiment, EMR was performed using augmented reality technique. We made an ablation of a targeted gastric (5 cm in diameter) without perforation and severe bleeding. Wounds after the resection were closed using a clipping device, which was directed by a combined effort from the right and left manipulators.

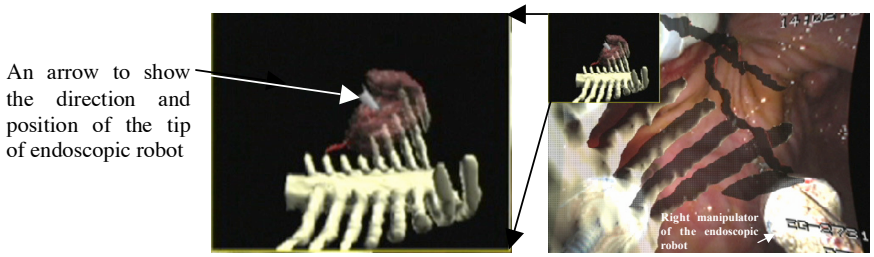


Fig. 9. A superimposed image on the GWS display. (Organ models (spine, ribs and hepatic artery) are superimposed onto the endoscope's image). Top left sub window shows the 3D location of the endoscope's tip in the coordinate system of the organ models.

3 Conclusions and Future

We described the process of the loading of the augmented reality functions for robotic surgery. There are many advantages of the super imposed 3D image for future developments of this kind of system. First, there is no need to display all the volume data fully reconstructed from an acquired CT or MR scan for the rapid and easy recognition of the operation conditions. If all of the volume data acquired from measurements were overlaid at the time of the surgical operation, it would create a confusing display for the surgeons. Instead, there needs to be effective logic available so that excess data can be deleted effectively. There also needs to be an innovative display method so that structures which the surgeon wants will be distinctive in the operating field. In addition, there is a need to develop 3D positioning measurements without using engineering markers that need to be sterilized every time. If we can overcome these issues, we will be able to more easily capture the inner structure of the organ as it changes during the operation. One of the most desired things in augmented reality is the realization of technology which can update whole changes of

elastic organs made during the operation. However, it is very difficult to imagine that this would be realized, even 10 years from now, unless there are significant advances in biological imaging technology. It is unlikely that using the smallest dose CT or highest speed MR scans will provide any advantages in this sense either. One major advance would be technology that can monitor in 4D space a marker free operation structure. If there are developments in this field accompanied with innovative changes in display function and 3D positioning measurements, navigation surgery will become more practical procedure and be closer to common clinical work.

References

1. Devernay F, Mourgues F, Coste-Maniere E. Towards endoscopic augmented reality for robotically assistedminimally invasive cardiac surgery. Proc. of the International Workshop on Medical Imaging and Augmented Reality (MIAR2001); 16-20.
2. Birkfellner W, Figl M, Huber K, Watzinger F, Wanschitz F, Hummel J, Hanel R, Greimel W, Homolka P, Ewers R, Bergmann H. A head-mounted operating binocular for augmented reality visualization in medicine - design and initial evaluation. IEEE Trans. Med. Imag. 2002; 21(8): 991-7.
3. Shahidi R, Bax MR, Maurer CR Jr, Johnson JA, Wilkinson EP, Wang B, West JB, Citardi MJ, Manwaring KH, Khadem R. Implementation, calibration and accuracy testing of an image-enhanced endoscopy system. IEEE Trans. Med. Imag. 2002; 21(12): 1524-35.
4. Hattori A, Suzuki N, Hashizume M, Akahoshi T, Konishi K, Yamaguchi S, Shimada M, Hayashibe M. A robotic surgery system (da Vinci) with image-guided function. Medicine Meets Virtual Reality 11 2003: 110-6.
5. Nicolau SA, Pennec X, Soler L, Ayache N. A Complete Augmented Reality GuidanceSystem for Liver Punctures: First Clinical Evaluation. MICCAI 2005, LNCS 3749, 539-47.
6. Giraldez JG, Caversaccio M, Pappas I, Kowal J, Rohrer U, Marti G, Baur C, Nolte LP, Ballester MAG. Design and clinical evaluation of an image-guided surgical microscope with an integrated tracking system. Int J CARS 2007; 1: 253-64.
7. Salisbury JK. The heart of microsurgery. Mechanical Engineering Magazine, ASME Int'l. 1998; 120(12): 47-51.
8. Reichenspurner H, Damiano RJ, Mack M, Boehm DH, Gulbins H, Detter C, Meiser B, Ellgass R, Reichart B. Use of the voice-controlled and computer-assisted surgical system ZEUS for endoscopic coronary artery bypass grafting. J Thorac Cardiovasc Surg 1999; 118: 11-6.
9. Guthart GS, Salisbury JK. The Intuitive Telesurgery System: Overview and Application. Proc. of the IEEE International Conference on Robotics and Automation (ICRA2000). San Francisco CA, April 2000.
10. Suzuki N, Sumiyama K, Hattori A, Ikeda K, Murakami EAY, Suzuki S, Hayashibe M, Otake Y, Tajiri H. Development of an endoscopic robotic system with two hands for various gastric tube surgeries. Medicine Meets Virtual Reality 11 2003: 349-53.

Refining the 3D surface of blood vessels from a reduced set of 2D DSA images

Erwan Kerrien[1], Marie-Odile Berger[1], and Jérémie Dequidt[2]

[1] Magrit project-team, INRIA Nancy-Grand Est/LORIA, France,
{Erwan.Kerrien,Marie-Odile.Berger}@loria.fr
[2] Alcove project-team, INRIA Lille-Nord Europe/LIFL, France,
Jeremie.Dequidt@lifl.fr

Abstract. Numerical simulations, such as blood flow or coil deployment in an intra-cranial aneurism, are very sensitive to the boundary conditions given by the surface of the vessel walls. Despite the undisputable high quality of 3D vascular imaging modalities, artifacts and noise still hamper the extraction of this surface with enough accuracy. Previous studies took the a priori that a homogeneous object was considered to make the reconstruction from the Xray images more robust. Here, an active surface approach is described, that does not depend on any particular image similarity criterion and grounds on high speed computation of the criterion derivatives. Mean square error and normalized cross-correlation are used to successfully demonstrate our algorithm on real images acquired on an anthropomorphic fantom. Preliminary results of coil deployment simulation are also given.

1 Introduction

An intracranial aneurism is a localized blood-filled dilation of a blood vessel that may occasionally rupture, causing hemorrhage, stroke or death. Medical treatment consists in excluding the aneurism from blood circulation either by surgical clipping or endovascular coil embolization. The latter, more recent, interventional technique is associated with an overall lower morbidity and mortality than surgery (9.5% at one year, compared to 12.2%). Furthermore, the recovery period following the intervention is short and very small aneurisms can be treated (down to 2 mm in diameter). However, the technical challenges that coil embolization faces remain important. It is believed that the still high recurrence rate after treatment would be reduced if the coil dimensions (loop diameter and length) were easier to choose, and coil placement and packing density could be better planned to ensure no blood circulates in the aneurism poach at the end of the treatment. This clearly emphasizes the need for simulation systems that can be used for physician training as well as for pre-operative planning.

Numerical simulations, such as blood flow or coil deployment in an intracranial aneurism, are very sensitive to the boundary conditions given by the surface of the vessel walls. This surface should be smooth to ensure the microtools glide

in a realistic way within the vasculature, and it should be accurate for the simulation to be as exact as possible. Despite the undisputable high quality of 3D vascular imaging modalities, artifacts and noise still hamper the extraction of this surface: blood flow artifacts in magnetic resonance angiography (MRA), tomographic artifacts in computed tomography angiography (CTA) and three-dimensional rotational angiography (3DRA), patient motion induced artifacts in all 3 modalities. Besides, digital subtracted angiography (DSA) remains the modality of reference when a very high precision is required, but it is limited to 2D images.

The work described in this paper aims at refining the surface of the vasculature extracted from any above 3D modality [1], so that this surface is accurate and smooth enough for subsequent simulations. The proposed approach is based on a deformable surface model whose internal force controls smoothness, and which evolves as an active surface until its Xray projections fit the DSA images. A steady streamline of works can be tracked back to the 90's [2] that aimed at making the tomographic reconstruction process more robust to noise and limited data by assuming that a finite set of homogeneous objects is to be reconstructed. Such works followed two main branches. A rather recent branch investigated the use of level-sets to model the object surface. However, this approach was only used for reconstructions from a dense number of views [3,4]. Most works model the object as a polyhedral surface, whose vertex positions are estimated in a Bayesian framework [5,6,7,8]. Free-form deformations [6,7] were also proposed to solve topological issues that could occur e.g. after collisions. This classical framework relies on the Mean Square Error (MSE) criterion, which is not always applicable in practice, since the raw density images are not readily available. Furthermore, optimization requires many derivations of the X-ray projection of the triangulated surface, which is time consuming. Both issues prevented any of these previous studies [6,7,8] from demonstrating their work on real images involving large surface models (thousands of voxels). Section 2 extends the classical framework to other iconic image similarity measures. Analysis of the X-ray projection then allows for insights to speed up the computation of derivatives. Such improvements enable us to demonstrate the global method onto real X-ray images, and give preliminary coil simulation results in section 3.

2 Method

2.1 Active surface model

The input of our refinement algorithm is a triangulated surface V of the vessels. This surface may have several connected components, all closed and oriented. Registered Xray images are assumed to be available, showing the object alone on a uniform background. Section 3 will give details on how these conditions are met in practice with vascular images.

The method we use originates in [9]. The mesh can be viewed as a dynamic system of particles that undergoes external forces F_{ext} and is constrained by

internal forces F_{int}. The Newtonian law of motion is applied on each vertex v_i:

$$m\ddot{v_i} + \gamma\dot{v_i} = F_{int}(v_i) + F_{ext}(v_i) \quad (1)$$

where m is the mass of the vertex and γ the damping (viscosity) factor. We used a simpler version of F_{int} than [9], namely $F_{int}(v_i) = \alpha(v_i - \bar{v_i})$, where $\bar{v_i}$ is the average of all neighbours of v_i and α a weight. This force relates to a curvature constraint.

We also set the common particle mass m to null. This has little impact on the convergence of the dynamic system. Then equation 1 leads to a linear system:

$$(A + \gamma I)V^{t+1} = F_{ext}(V) + \gamma V^t \quad (2)$$

where $A + \gamma I$ is a sparse matrix which is iteratively inverted.

When image data are 3D, external forces are classically expressed as the 3D gradient of the image, which comes from the derivative of an energy expressed as the square gradient norm. Similarly, we consider an external energy term based on an image similarity criterion C (to be minimized):

$$F_{ext}(v_i) = \beta \sum_{D \in \mathcal{D}} \frac{\partial C(\mathbb{P}_D(V), D)}{\partial v_i}$$

where $\mathcal{D}$ is the set of Xray images and $\mathbb{P}_D(V)$ is the Xray projection of the surface under the same incidence as Xray image D, and β a weight. C is traditionnally the Mean Square Error (MSE) [7,6,8]: $C = \sum_D \|\mathbb{P}_D(V) - D\|^2 / N_p$ (N_p is the number of pixels). We shall also consider in the following experiments the Normalized Centered Cross-correlation (NCC):

$$C = 1 - \sum_D E[\mathbb{P}'_D(V)D']/(\sigma(\mathbb{P}'_D(V))\sigma(D'))$$

with E the expectancy and, for a given image I, $\bar{I}$ its mean, $I' = I - \bar{I}$, the centered image and $\sigma(I')$ its standard deviation.

2.2 Computation of the external force - Surface Xray projection

Computing the external force implies computing the derivatives of C with respect to each v_i. This computation is classically done numerically by using the central difference approximation:

$$\frac{\partial C(\mathbb{P}_D(V), D)}{\partial v_{i,j}} \approx \frac{C(\mathbb{P}_D(V + \delta_{i,j}), D) - C(\mathbb{P}_D(V - \delta_{i,j}), D)}{2\delta}$$

where $v_{i,j}$ is the j^{th} component ($j \in [1,3]$) of vertex v_i and $V + \delta_{i,j}$ (resp. $V - \delta_{i,j}$) is the surface V perturbed by adding (resp. subtracting) δ to $v_{i,j}$.

This requires computing many Xray projections of the surface, which is time consuming. Soussen [8] derived analytic expressions of the derivative of $\mathbb{P}_D(V)$ but only in the case of parallel projection. This result is not applicable to our case, since DSA images are acquired under central projection geometry. Nevertheless, we demonstrate here how to speed up the Xray projection of a triangulated surface and perform fast numerical computation of the external force.

Xray projection of the surface The triangulated surface V is composed of a finite set of connected components. Each component is closed and oriented. We further assume in what follows that the surface is a topological open: a ray tangent to the surface will be considered as passing outside the surface, so that a ray intersects a triangle at 0 or 1 point.

For a given pixel (u, v), the ray from that pixel to the focal point O_D intersects the surface at an even number of points $(P_k)_{k\in[0,2M-1]}$ such that the ray enters the surface at P_{2i} and goes out of the surface at P_{2i+1}. We say that P_{2i} is an *entry point* and P_{2i+1} is an *exit point* (see fig. 1, left). The Xray projection of the surface is then given by: $\mathbb{P}_D(V)(u, v) = \sum_{i=0}^{M-1} \|P_{2i+1} - P_{2i}\|$ Since all the intersection points are aligned with O_D, this can be rewritten as:

$$\mathbb{P}_D(V)(u, v) = \sum_{k=0}^{2M-1} \sigma_k \|P_k - O_D\|$$

where $\sigma_k = -1$ if P_k is an entry point and $\sigma_k = 1$ if P_k is an exit point. Each intersection point and thereby each triangle can thus be processed independently. Furthermore, The σ coefficient will be the same for all points on a given triangle t, depending only on the way t is seen from the focal point: Each triangle t can therefore be qualified as either entry, exit or tangent triangle depending on $\sigma(t) = \text{sign}(n_t^T(\bar{t}-O_D))$, with n_t the normal to t (exponent T is for transposition) and $\bar{t}$ its center of gravity. As a result, the Xray projection can be expressed as:

$$\mathbb{P}_D(V) = \sum_{t\in V} \sigma(t)\mathbb{P}(t) \tag{3}$$

Special cases A ray may occasionnally stab an edge or a vertex of a triangle. For the edge case (fig. 1, middle), if both triangles sharing the edge are entry (resp. exit) triangles, the intersecting point is an entry (resp. exit) point. Otherwise, the ray is tangent (no intersection). For the vertex case (fig. 1, right), all triangles sharing this vertex (so-called the *vertex fan*) are projected on the image plane. Each triangle projects onto another triangle. The sum of the oriented angles at the common summit in the projected fan is either 2π for an entry point, -2π for an exit point or 0 in the tangent case.

Degenerate cases occur when one of the involved triangles is tangent to the ray. The full triangle case is trivial since no intersection occur. The edge and vertex cases are resolved using Simulation of Simplicity (SoS) [10]: one (edge case) or two (vertex case) vertices of the tangent triangle are moved by $-\epsilon n_t$ (with $\epsilon > 0$). The status of the intersecting point can then be set and does not depend on ϵ (the proof is too long to hold on this paper).

Computation of the external forces Computing the external forces implies updating the Xray projection of the surface after a single vertex v_i was moved by δv. The motion only changes the vertex fan of v_i. Following equation 3, this update can be made by subtracting the projection of the vertex fan of v_i, moving

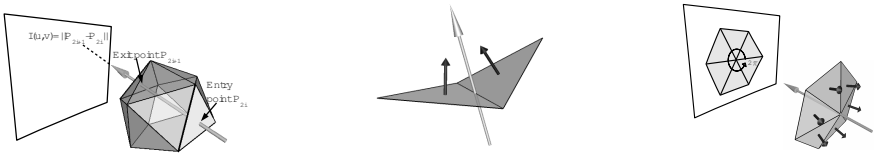


Fig. 1. Xray projection: (left) a ray intersects the surface at an even number of points: each entry point corresponds to an exit point. (middle) An exit edge connects 2 exit triangles; (right) entry vertex: the sum of the angles on the projected fan is 2π. (White arrow: Xray direction, black arrows: outward face normals.)

v_i and adding the projection of the new vertex fan:

$$\mathbb{P}_D(V') = \mathbb{P}_D(V) - \sum_{t \in \mathcal{V}_i} \sigma(t)\mathbb{P}_D(t) + \sum_{t' \in \mathcal{V}'_i} \sigma(t')\mathbb{P}_D(t')$$

where $\mathcal{V}_i$ is the vertex fan of v_i, $\mathcal{V}'_i$ is the new vertex fan and V' the surface after the vertex motion. This computation only involves a few triangles and concerns a few pixels in the image. The update of the similarity criterion is therefore very fast, leading to a fast computation of the central difference scheme.

3 Results

The whole algorithm was implemented in C++ based on the open source Gnu Triangulated Surface library (GTS) (`http://gts.sourceforge.net`). For all the experimentations, the following common configuration was chosen: the surface was retriangulated each 10 iterations; for the numerical derivation $\delta = 2$. The only varying parameters were: number of iterations; internal coefficient α ; external coefficient β; and viscosity parameter γ.

3.1 Synthetic ellipsoid with MSE

A first synthetic experiment was conducted, taking an ellipsoid with a much larger main axis than the other two (equal) axes, so as to approach the shape of a blood vessel. 3 Xray views were simulated: antero-posterior (AP), lateral and top. The ellipsoid was slightly turned so that its main axis was not exactly aligned with the lateral view direction. The process was initialized with a much thinner ellipse. Indeed, active surfaces naturally tend to shrink. Looking for an enlargement should show up the effects of the external force. 500 iterations were made, with $\alpha = 1$ $\gamma = 10$ and $\beta = 100$. The MSE similarity criterion was chosen. Figure 2 shows some steps in the optimization: the convergence was fast for the occluding vertices, as in classical computer vision cases, but the Xray density information still acted, yet much more slowly, to make the other vertices converge. The shortening of the ellipsoid is to be noted: this is due to much higher internal forces at the extremities where the curvature is very strong.

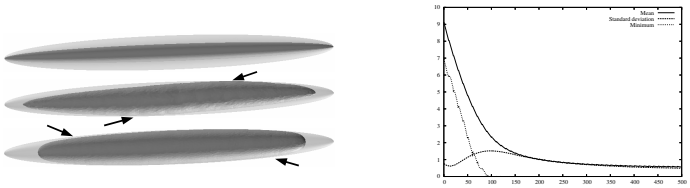


Fig. 2. Synthetic experiments: (left) Screen captures at iteration 0, 100 and 400. The true shape is translucent. The shape first evolves to fit the occluding contours (arrows on the middle image), and then slowly converges on parts where only density information is available. (right) evolution of the distance to the true shape.

3.2 Anthropomorphic phantom

A second experiment was performed on real images of an anthropomorphic phantom of the brain vasculature (Elastrat, Geneva: rigid phantom with 3 aneurisms).

Surface data A 3DRA was acquired on the phantom on an Innova 4100 angiography machine (GE Healthcare: 512^3 voxel data cube, 0.25 mm voxel size) and the vessel isosurface was extracted using the marching cube from the GTS. In the absence of a CAD model, this surface was considered as ground truth. The initial surface was obtained by simulating MRA conditions: the 3DRA volume was subsampled to a voxel size of 1 mm, using nearest neighbour interpolation, and an isosurface extracted using marching cube (10210 vertices).

DSA images 4 DSA sequences were available for the phantom: one AP and one lateral view, as well as two other "working" views, that the physician would have used to treat the aneurism. The bolus density 3D map can be considered constant in the vessels, if averaged over a sufficient time (e.g. arterial time). Due to the linearity of the projection operator, this volume projection is the average of the first DSA images in the sequence: the average over the first 10 images was considered as Xray data. However, since equality cannot be assumed between the real Xray data and the synthetic surface projection, NCC was considered in this experiment. DSA images were registered with the 3DRA volume using [11]. Only crops of about 200×130 pixels around the aneurism were considered to speed up the process. As a result, the optimization concerned 5900 vertices.

Experimentations The algorithm was run in two passes of 100 iterations each: a first step with a very small internal force ($\alpha = 0.001, \gamma = 30, \beta = 60000$) to correct for the larger differences, but preserve the small vessels and a second step to refine the surface and smooth it ($\alpha = 0.03, \gamma = 2, \beta = 10000$). Figure 3 shows the result of the reconstruction: the shape of the aneurism and the largest vessels are well recovered and the surface is even smoother than in the ground truth image. The small vessel surface is however more irregular. Indeed, the internal force is much stronger on small vessels with highly curved cross-section and

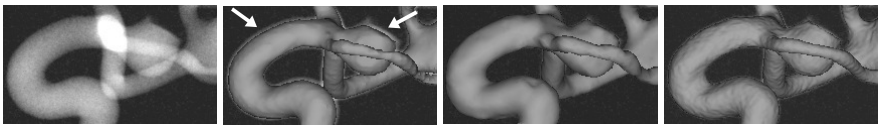


Fig. 3. Experiments on real data: (from left to right) lateral Xray DSA image; initialization: notice the error at the top of the aneurism and vessel boundaries (arrows); final surface: the aneurism surface and the large vessels were recovered, and the surface is smoother than in the ground truth surface; ground truth surface: notice the size of the small vessel before the aneurism is underestimated compared to the DSA image.

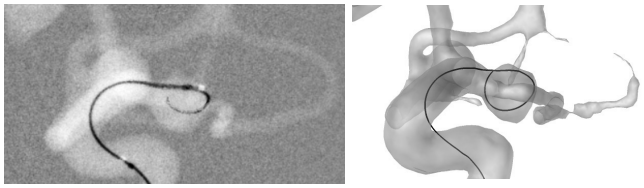


Fig. 4. Simulation of coil deployment: (left) actual coil position, as seen in DSA; (right) simulation of the coil within the vessel surface given by our algorithm.

should be adapted to the local vessel size. 47 minutes were necessary for the 200 iterations. As a comparison, extrapolating Soussen's results [8] to our experiment would require 1470 minutes (350 seconds for 10 iterations, 210 vertices and 9 64×64 images, in parallel projection).

3.3 Simulation of coil deployment

A preliminary experiment to simulate a coil deployment was made using the simulation method presented in [12]. The vessel surface was segmented and refined using our algorithm (see fig. 3). Figure 4 shows the actual coil seen in a DSA image and compares it with the simulated coil within the vessels in a similar orientation to demonstrate a very encouraging realism of the simulation (note that some vessels shrunk because they were not visible in the cropped DSA images).

4 Discussion and future works

Despite tremendous improvements in 3D vascular medical imaging, discrepancies still exist with DSA images that the interventional radiologist uses as feedback to his/her surgical gesture. Such differences could be reduced by a binary object reconstruction method such as [8], let apart 3 main issues that prevent such methods from being applicable on real data: good initialization, since the problem is very ill-posed; appropriate image similary criterion, since the input images are in general known up to a look-up transform; fast computation of the criterion

derivatives, since many vertices are involved in real datasets and many iterations are required.

The work described in this paper proposes to initialize the surface from a 3D vascular imaging modality, extends previous reconstruction works to any image similarity criterion, and describes how derivative computations can be sped up. While previous works were only demonstrated on synthetic data, results are here provided on real datasets of an anthropomorphic phantom. As future work, parameter tuning is not a trivial task, that we expect will be better circumscribed after a more extensive testing on actual patient data. Furthermore, Lachaud's work [9] will be integrated to enable topological changes on the surface, that may be necessary to separate two very close vessels.

As an illustrative example application, preliminary, though very realistic, simulation results are provided for the deployment of a coil within an intracranial aneurism. Other applications can exploit our method, when the 3D vessel needs to updated from a DSA sequence: estimation/correction of blood vessel deformation when the micro-tools are inserted or due to the arterial pulsatility, or blood flow reconstruction in 3D by a 3D bolus tracking approach.

References

1. Kirbas, C., Quek, F.: Vessel extraction techniques and algorithms: A survey. In: Proc. of the 3rd Symp. on Bioinformatics and Bioengineering. (2003) 238–45
2. Bresler, Y., Fessler, J.A., Macovski, A.: A Bayesian approach to reconstruction from incomplete projections of a multiple object 3D domain. IEEE Trans. PAMI **11**(8) (1989) 840–58
3. Bruandet, J.P., Peyrin, F., Dinten, J.M., Barlaud, M.: 3D tomographic reconstruction of binary images from cone beam projections: A fast level set approach. In: IEEE Int. Symp. on Biomedical Imaging - ISBI'02. (2002) 677–80
4. Feng, H., Karl, W.C., Castañon, D.A.: A curve evolution approach to object-based tomographic reconstruction. IEEE Trans. Image Processing **12**(1) (2003) 44–57
5. Milanfar, P., Karl, W.C., Willsky, A.S.: Reconstructing binary polygonal objects from projections: A statistical view. CVGIP: Graphical Models and Image Processing **56**(5) (1994) 371–91
6. Hanson, K.M., Bilisoly, R.L., Cunningham, G.S.: Kinky tomographic reconstruction. In: SPIE Medical Imaging: Image Processing. Volume 2710. (1996) 156–66
7. Battle, X.L., Le Rest, C., Turzo, A., Bizais, Y.: Three-dimensional attenuation map reconstruction using geometrical models and free-form deformations. IEEE Trans. Med. Imag. **19**(5) (2000) 404–11
8. Soussen, C., Mohammad-Djafari, A.: Polygonal and polyhedral contour reconstruction in computed tomography. IEEE Trans. Image Processing **13**(11) (2004) 1507–23
9. Lachaud, J.O., Montanvert, A.: Deformable meshes with automated topology changes for coarse-to-fine 3D surface extraction. Medical Image Analysis **3**(2) (1999) 187–207
10. Edelsbrunner, H., Mücke, E.P.: Simulation of Simplicity: A technique to cope with degenerate cases in geometic algorithms. ACM Trans. Graphics **9**(1) (1990) 66–104
11. Kerrien, E., Berger, M.O., Maurincomme, E., et al.: Fully automatic 3D/2D subtracted angiography registration. In Taylor, C., Colchester, A., eds.: Proceedings of

the Second International Conference on Medical Image Computing and Computer-Assisted Intervention, Cambridge (UK). Number 1679 in Lecture Notes in Computer Science, Springer Verlag (September 1999) 664–671
12. Dequidt, J., Marchal, M., Duriez, C., Cotin, S.: Interactive simulation of embolization coils: Modeling and experimental validation. In: Proceedings of the 11th International Conference on Medical Image Computing and Computer Assisted Intervention, New York. Lecture Notes in Computer Science, Springer Verlag (September 2008) Accepted for publication.

Contextually Enhanced 3D Visualization for Multi-burn Tumor Ablation Guidance

Andrei State[a,b], Hua Yang (PhD)[a,1],
Henry Fuchs (PhD)[a], Sang Woo Lee[a], Patrick McNeillie[d], Charles Burke (MD)[c]

Departments of [a]Computer Science and [c]Radiology, and [d]School of Medicine, University of North Carolina at Chapel Hill; [b]InnerOptic Technology, Inc. (all in Chapel Hill, NC, USA)
andrei@cs.unc.edu,hua.yang@kitware.com,
{fuchs,leejswo}@cs.unc.edu,{patrick_mcneillie,charles_burke}@med.unc.edu

Abstract. We introduce visualization methods for augmented and virtual reality displays that guide minimally invasive tumor-targeting interventions, such as radio-frequency ablation (RFA). Our experimental RFA guidance system performs virtual ablations that geometrically match the real ablations in statically registered phantoms. It provides a head-tracked stereoscopic display that shows a real-time, motion-tracked ultrasound scan, the avatar of a motion-tracked RFA probe, and a registered tumor, which is incrementally "carved" away with each successive ablation pass. An analog bar graph displays tentative ablation coverage fractions for tumor and healthy tissue in real time. We introduce our system, describe the visualization techniques, and present initial results.

1 Introduction and Motivation

Minimally invasive techniques are commonly used to treat a variety of cancerous lesions. The interventions include ablative methods such as cryo-ablation, microwave ablation, and radio-frequency ablation (RFA). These are characterized by the need to place percutaneously a needle-like instrument ("probe") in a precise location within the tumor, usually guided by real-time 2D ultrasound imaging and/or pre-operative scans such as computer tomography (CT) or magnetic resonance imaging (MRI). Here we focus on hepatic tumor RFA [1], the driving application in our work.

Ultrasound offers real-time visualization while advancing an RFA probe into the lesion, which may provide more flexibility than either CT or MRI in terms of approach angles. However, conventional 2D ultrasound presents only planar cross-sections of the scanned tissue, while tumors are three-dimensional and can have complex shapes. Thus, even for the most experienced physician, it can be difficult to position the RFA probe optimally to maximize the tumor ablation while minimizing collateral injury. When treating larger lesions, adjunctive techniques may be required for complete ablation of tumor cells. One strategy that is commonly employed is the use of multiple, overlapping (successive or simultaneous) ablations ("multi-burn"). Mathematical models have been developed to optimize this strategy [2] and illustrate its inherent difficulties. In addition, the ablation zone often is rendered obscure on the

[1] now at Kitware, Inc., Carrboro, North Carolina, USA

ultrasound image due to the release of nitrogen gas [3]. We expect therefore that an operator who views real-time, intra-operative ultrasound data that is annotated with registered pre-operative data (showing a 3D reconstruction of the target lesion) will optimally position the RFA probe and will perform RFA more efficiently and accurately than under conventional ultrasound guidance alone. Such a "contextually enhanced" 3D display can also show a dynamic, incrementally updated visualization of the tumor as it is being ablated.

In this paper, we describe first steps towards creating such a visualization for an RFA guidance system, which is currently based on 2D ultrasound and pre-operative MRIs. Our techniques are suitable for display methods with enhanced depth perception, preferably using stereoscopy and head-motion parallax, such as augmented reality (AR) see-through head-mounted displays (STHMDs), or head-tracked, "fish-tank" virtual reality (HTVR) systems [4]. After surveying relevant previous work, we give an overview of our guidance system and introduce the multi-burn visualizations. With the help of a newly developed, ablatable liver phantom, we demonstrate the visualization's potential for mirroring a real ablation procedure as it guides the physician through it. We conclude with suggestions for future investigations.

2 Previous Work

Medical instrument guidance is addressed by several commercial systems. These typically use optoelectronic or magnetic motion tracking for interventional tools. The tools' poses or trajectories are displayed in real time and superimposed on pre-operative scans, usually in 2D cross-sections. Manufacturers include *BrainLab, Medtronic, Stryker, Veran Medical* and *Ultraguide* (no longer operating). All commercial systems we know of use conventional 2D monitors as displays; some (*BrainLab, Stryker*) present perspective views of reconstructed patient data (3D scans) together with instrument pose. No commercial guidance systems we are aware of use stereoscopic and/or head-tracked displays, and we are also unaware of any commercial guidance systems that might be employing the type of visualization we propose here.

On the other hand, surgery simulation systems [5] incorporate virtual cutting [6] visualizations that are analogous to the methods we present below—with the difference that they are used in simulation, not interventional guidance. In contrast to the above-mentioned commercial guidance systems, surgery simulators may use displays with stereoscopic imaging and head-motion parallax. The same applies to guidance system research prototypes, e.g., the STHMD-based systems developed by Nassir Navab's team [7], as well as our own STHMD-based prototypes (referenced below).

Other relevant work includes an experimental AR guidance system [8] that combines real-time 2D ultrasound scans with segmented data from pre-operative CT scans (e.g., vessels). The developers report positive feedback from two experienced interventionists recruited to guide an RFA probe into a phantom using their system.

At the University of North Carolina, a team led by Elizabeth Bullitt has developed a technology for registering 3D vessel trees from pre-op scans with intra-op angiograms. This technique is integrated into a real-time guidance system for transthoracic intrahepatic portosystemic shunt (TIPS) placement, a difficult endovascular in-

tervention. The TIPS project developed a head-tracked stereoscopic visualization system [9] that is similar to ours.

3 Guidance System Overview

We have been developing STHMD-based guidance systems aimed at minimally invasive interventions [10][11][12][13]. For this work, we have preferred an HTVR-based platform for a number of application-dependent reasons; these are discussed in an earlier paper [14], which also contains a brief description of our guidance system. We give a full overview here, with the understanding that our methods can also be applied to STHMDs (Fig. 1) or other AR devices, for example through-the-camera displays.

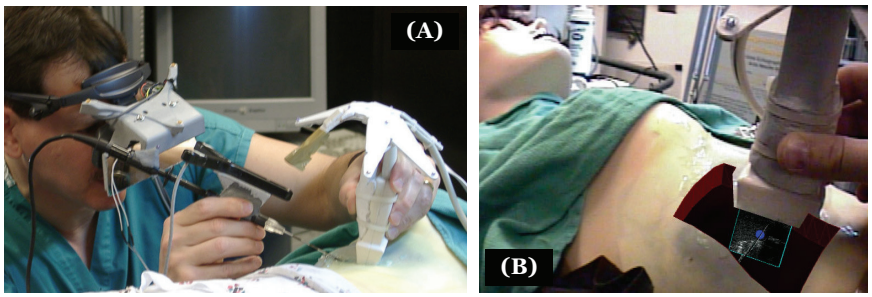


Fig. 1. (A) Our early prototype AR guidance system for liver RFA with video see-through HMD and motion-tracked tools (ultrasound transducer, RFA probe). (B) View through HMD: The ultrasound scan is below the transducer, within a synthetic opening into the phantom.

Our current HTVR-based prototype guidance system (Fig. 2A and 2B) targets liver RFA procedures. It integrates a 2D ultrasound scanner, a stereoscopic display (*Planar StereoMirror*™), an optoelectronic motion tracker (*NDI Optotrak Certus*) and a personal computer; it presents to the head-tracked physician (typically an interventional radiologist) a three-dimensionally merged, stereoscopic virtual environment (Fig. 2C) that represents an area within the patient. Within the display, the physician views a correctly scaled and dynamically updated ultrasound image that moves as if it were attached to the (schematically represented) handheld transducer; the physician also sees a dynamic representation (an avatar) of the RFA probe, which includes the (in reality invisible) part of the probe that is inside the patient. The estimated burn area of the RFA probe is shown as a wireframe surface (yellow sphere in Fig. 2C). We also calculate and display as a closed curve (not visible in these images) the burn area's intersection with the ultrasound slice. A set of 3D-guidance lines, visible at right in Fig. 2C, continually informs the physician about the current spatial relationship between RFA probe and ultrasound slice. This "contextually enhanced" view of the intervention scenario facilitates targeting using natural hand-eye coordination.

In contrast to STHMDs, the HTVR display does not provide registration between the computer-enhanced environment and the patient; however, we operate it with "orientation alignment," analogous to the "orbital mode" for HMDs [15], leaving only

a translational offset between 3D display contents and patient internals. An earlier publication [14] provides more detail and also describes other modalities we use to map user-induced instrument translation to the fixed 3D display. Our head tracking and eye calibration techniques for HTVR-type displays are described elsewhere [16].

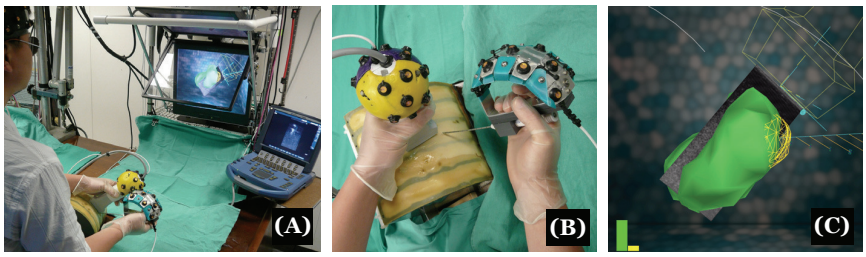


Fig. 2. (A) Guidance system for liver tumor RFA, in use on a liver phantom. (B) Motion-tracked ultrasound transducer and RFA probe. (C) The guidance system shows the pre-segmented tumor (green) within the phantom, registered with the real-time ultrasound scan, as well as the RFA probe and its 3D burn area (yellow wireframe sphere) intersecting the tumor. The bar graph at lower left shows the ratio of tumor (green) vs. healthy tissue (yellow) within the 3D burn area corresponding to the current probe placement.

4 Multi-burn Guidance Techniques

Virtual Ablation. In an initial multi-burn experiment (Fig. 3A and 3B), we introduced *virtual ablation*: by taking advantage of the known (tracked) position and orientation of the RFA probe, the system can "carve" out in the 3D display the area ablated from the target tumor, as the actual ablation takes place within the phantom. The virtual representation of the RFA probe within the 3D display thus acts as an "ice cream scoop," successively removing tumor fragments as they are being ablated (Fig. 3C)—provided the virtual ablation is initiated whenever a real ablation takes place, and provided that the exact shape of the RFA probe's burn area is known. We expect this technique to become an indispensable "blind navigation" aid for successive overlapping burns, given that, as already mentioned, the real-time ultrasound scan shows ever more artifacts with each successive burn, and so its utility diminishes rapidly. In contrast, our method continually tracks the RFA probe and virtually removes ablated tumor material, providing a clear visualization of un-ablated tumor tissue that is to be targeted for further treatment. For example, in Fig. 3C, the spherical tumor was virtually ablated several times.

The "carving" technique is currently implemented in MATLAB and uses volume transformation and subtraction calculations coupled with marching-cubes surface extraction. Because of that, each virtual ablation takes several seconds to complete. By switching to volumetric hardware-assisted rendering, we could eliminate the need for surface extraction and take advantage of parallelism within graphics processor units (GPUs). This would accelerate the virtual ablation to near-real-time speeds, which at first seems unnecessary given that the physical ablation takes minutes to complete. However, consider the utility of seeing in real time the potential effect of an ablation

triggered at the current probe position in real time, while manually attempting to move the probe to the most suitable location for the next ablation pass.

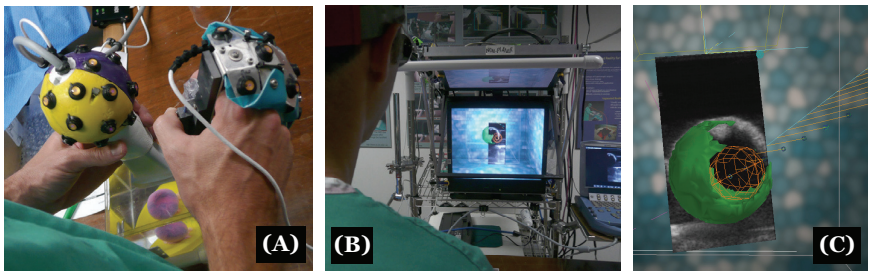


Fig. 3. (A) Simple phantom for initial multi-burn guidance experimentation consists of a Play-doh sphere affixed to the bottom of a water bath. (B) Using the multi-burn visualization technique in the guidance system. (C) Guidance system display during virtual ablation of the tumor with successive burns; after each burn, the ablated tumor area is "carved" away, showing only the remnant tumor.

Ablation Fractions Graph. The second enhancement is an analog bar graph display (lower left in Fig. 2C) that shows—in real time—the fractions of tumor (green) and of healthy tissue (yellow) that would be ablated if ablation were triggered at the current probe position. ("Fraction" implies a number derived from a scalar voxel count.) The additional calculations for this display component are easily supported by the data processing required for the carving visualization. The simple, but highly informative display should provide additional help for ablating as much tumor as possible, and as much surrounding healthy tissue as necessary for low recurrence probability. It is useful even for single-burn procedures, where it can confirm that the entire tumor is contained within the tentative ablation zone.

5 Matching Real and Virtual Ablation

In order to evaluate the new technique, we constructed a phantom containing ablatable tumor and background materials (Fig. 4A and 4B), based on acrylamide-albumin chemistry [17][18]. The phantom images realistically under ultrasound as well as MR, and allows segmentation of four tissue types: ablated and un-ablated tumor, as well as ablated and un-ablated background (Fig. 4C). This characteristic should enable systematic measurements of the performance of guidance systems such as ours. As Fig. 4A shows, we mounted water-filled tubes on the phantom to aid with (static) registration.

Using ITK-SNAP software, we segmented from the phantom's pre-RFA MRI the tumor (green in Fig. 4B) and the two water-filled tubes (blue in Fig. 4B). Before ablation, we calibrated the physical tubes with the Certus' measurement pointer. We obtained static registration between the phantom's pre-operative tumor surface and the real-time, motion-tracked intra-operative 2D ultrasound scans (Fig. 5A) by transforming the MR data such that the MR-segmented tubes coincide with the physical tubes.

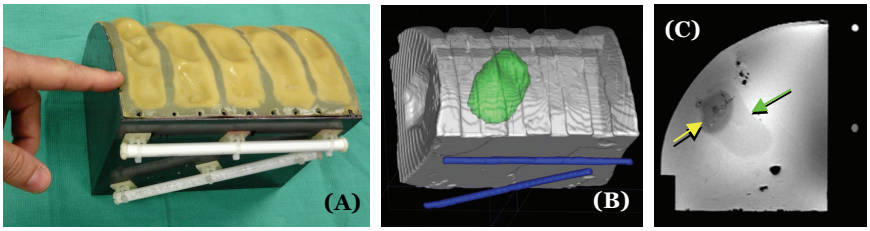


Fig. 4. (A) Liver phantom for multi-burn experimentation, equipped with (white) water-filled tubes designed to register the pre-segmented tumor—see pre-RFA MR data in (B), with tubes shown in blue—with the physical phantom (A). The post-RFA MR section (C) shows the elongated tumor (green arrow) and the intersecting burn area (yellow arrow, cf. Fig. 5B and 5C).

Using the guidance system as described above, we placed the RFA probe with the goal of ablating a fragment of the tumor within the phantom (Fig. 2C). Virtual ablation was triggered as the RFA probe was energized. The remnant tumor calculated by virtual ablation is shown in Fig. 5B. After ablation, we re-imaged the phantom and segmented the phantom's remnant tumor from the post-MRI data (Fig. 4C). The result of the segmentation (Fig. 5C) closely resembles the virtual ablation. The remnant virtual and segmented tumors (Figs. 5B and 5C) are similar in shape, and both the virtual and MR-segmented ablation volumes contain approximately 15 cm^3 (slightly less than a cubic inch). This represents an initial confirmation that our method can be applied to multi-burn guidance, since we can expect that with accurate tracking and registration, as well as with exact calibration of the RFA probe's burn area, the virtual and real ablations will not "drift" too far apart after multiple successive burns.

Fig. 5. (A) Accurate registration between pre-segmented tumor (green wireframe) and real-time ultrasound scan. The partially ablated tumor calculated by the guidance system (B) approximately matches the remnant tumor (C) segmented from the post-ablation MRI.

6 Conclusions and Future Work

We have demonstrated intuitive visualization methods suitable for guidance of multi-burn ablation procedures. As a result of accurate instrument tracking, and by using a realistically imaging, registered phantom, we can perform virtual ablations "in sync"

with real ones, as compactly demonstrated by the companion images in Fig. 5B and 5C. The new visualization techniques (three-dimensional virtual ablation and dynamic ablation fractions graph) were integrated into an experimental liver RFA guidance system that uses a head-tracked fish tank VR stereoscopic display; they are equally suitable for AR systems employing see-through HMDs, as well as for other 3D display modalities. Furthermore, they can easily be adapted for simultaneous ablations (which use multiple RFA probes energized at the same time), where they can guide pre-RFA probe placement such that the multiple, overlapping burn areas encompass all tumor tissue.

We can currently use these methods only on motionless, rigidly registered phantoms. To maintain registration with a moving, breathing patient, our rigid, static pre-operative registration is insufficient. Other researchers have used methods such as skin-mounted fiducials [19] to track patient motion. But for highest possible accuracy, one has to "attach" the pre-operative data to the tumor tissue itself, or rather, given that real-time ultrasonic imaging in the ablation area degrades after the first ablation, to the tumor's immediate neighborhood. Real-time 3D ultrasound echography (or "4D ultrasound" by its commercial name) can image the entire surroundings of a tumor and is therefore more suitable for dynamic intra-op/pre-op registration than its "ambiguously two-dimensional" counterpart. We are considering techniques such as vessel-based feature-to-image registration [20], which can match pre-operative MR imagery to intra-operative 3D ultrasound. This method can be extended to non-rigid deformations and has the potential of eventually operating in real time (in a GPU-based implementation), as an integral component of a future guidance system. We expect that as tracking, miniature-display and -camera technologies advance, such systems will ultimately be based on see-through head-mounted displays.

Acknowledgments

This work was funded by the National Institutes of Health (1 R01 CA101186-01A2). John E. Thomas and Kurtis Keller provided mechanical engineering support. We also thank Stephen Aylward, John Cullen (VMD, PhD), Sashi Gadi, Matthew Mauro (MD), Sharif Razzaque (PhD), Herman Towles and Ginny Turner. The guidance system's software incorporates *libwave* by Dave Pape (ported to Windows by us).

References

1. Hyunchul Rhim, MD, S. Nahum Goldberg, MD, Gerald D. Dodd, III, MD, Luigi Solbiati, MD, Hyo Keun Lim, MD, Massimo Tonolini, MD and On Koo Cho, MD: Essential Techniques for Successful Radio-frequency Thermal Ablation of Malignant Hepatic Tumors. Radiographics 2001 Oct;21 Spec No:S17-S35, S36-9.
2. Y. S. Khajanchee, D. Streeter, L. L. Swanstrom and P. D. Hansen: A mathematical model for preoperative planning of radiofrequency ablation of hepatic tumors. Surgical Endoscopy 2004 Apr;18(4):696-701.

3. Steven Raman, David Lu, Darko Vodopich, James Sayre and Charles Lassman: Creation of radiofrequency lesions in a porcine model: correlation with sonography, CT, and histopathology. AJR American J. of Roentgenology Nov 2000; 175(5):1253-1258.
4. Michael Deering: High Resolution Virtual Reality. Proc. SIGGRAPH '92. Computer Graphics, 26(2), 195-202.
5. U. Meier, O. López, C. Monserrat, M.C. Juan and M. Alcañiz: Real-time deformable models for surgery simulation: a survey. Computer Methods and Programs in Biomedicine, 2005 Mar;77(3):183-97.
6. C. Bruyns,, S. Senger, A. Menon, K. Montgomery, S. Wildermuth, R. Boyle: A survey of interactive mesh cutting techniques and a new method for implementing generalized interactive mesh cutting using virtual tools. J. of Visualization and Comp. Anim., 13(1), 21–42.
7. Christoph Bichlmeier, Sandro Michael Heining, Mohammad Rustaee, Nassir Navab: Virtually Extended Surgical Drilling Device: Virtual Mirror for Navigated Spine Surgery. Proc. MICCAI 2007, 434-441.
8. Thomas Stüdeli, Denis Kalkofen, Adinda Freudenthal: How do Interventionists navigate to the tumour? Proc. 39th Nordic Ergonomics Society (NES'07).
9. D. Maupu, M.H. VanHorn, S. Weeks, E. Bullitt: 3D stereo interactive medical visualization. IEEE Comput Graph Appl. 2005, 25(5):67-71.
10. Andrei State, Mark A. Livingston, William F. Garrett, Gentaro Hirota, Mary C. Whitton, Etta D. Pisano (MD) and Henry Fuchs: Technologies for Augmented-Reality Systems: Realizing Ultrasound-Guided Needle Biopsies. Proc. SIGGRAPH '96. Computer Graphics Proceedings, Annual Conference Series, 1996, ACM SIGGRAPH, 439-446.
11. Michael Rosenthal, Andrei State, Joohi Lee, Gentaro Hirota, Jeremy Ackerman, Kurtis Keller, Etta D. Pisano, Michael Jiroutek, Keith Muller and Henry Fuchs: Augmented reality guidance for needle biopsies: An initial randomized, controlled trial in phantoms. Medical Image Analysis Sep 2002; 6(3):313-320.
12. Andrei State, Kurtis Keller, Michael Rosenthal, Hua Yang, Jeremy Ackerman and Henry Fuchs: Stereo imagery from the UNC augmented reality system for breast biopsy guidance. Studies in Health Technology and Informatics. 2003; 94:325-328.
13. Andrei State, Kurtis P. Keller and Henry Fuchs: Simulation-Based Design and Rapid Prototyping of a Parallax-Free, Orthoscopic Video See-Through Head-Mounted Display. Proc. International Symposium on Mixed and Augmented Reality (ISMAR) 2005, 28-31.
14. Henry Fuchs, Andrei State, Hua Yang, Tabitha Peck, Sang Woo Lee, Michael Rosenthal (MD/PhD), Anna Bulysheva, Charles Burke (MD): Optimizing a Head-Tracked Stereo Display System to Guide Hepatic Tumor Ablation. Studies in Health Technology and Informatics. 2008; 132:126-131.
15. James C. Chung: A comparison of head-tracked and non-head-tracked steering modes in the targeting of radiotherapy treatment beams. Proc. 1992 Symposium on Interactive 3D Graphics, 193-196.
16. Andrei State: Exact Eye Contact with Virtual Humans. Proc. IEEE International Workshop on Human Computer Interaction 2007, LNCS, vol. 4796:138-145.
17. Mark McDonald, Shanna Lochhead, Rajiv Chopra and Michael Bronskill: Multi-modality tissue-mimicking phantom for thermal therapy. Physics in Medicine and Biology 2004; 49:2767-2778.
18. Patrick McNeillie: Two compartment phantom for modeling hepatic tumors. Presentation at the Institute of Biological Engineering 2008 Annual Meeting, March 7, 2008, Research Triangle Park, North Carolina.
19. Stéphane Nicolau: Un système de réalité augmentée pour guider les opérations du foie en radiologie interventionnelle. Doctorate thesis, Université de Nice – Sophia Antipolis, 2004.
20. S. Aylward, J. Jomier, S. Weeks and E. Bullitt: Registration and Analysis of Vascular Images. International Journal of Computer Vision, Nov 2003; 55(2-3):123-138.

Video to CT Registration for Image Overlay on Solid Organs

Balazs Vagvolgyi, Li-Ming Su, Russell Taylor, and Gregory D. Hager

Center for Computer-Integrated Surgical Systems and Technology
Johns Hopkins University
Baltimore, MD 21218
{vagvoba, hager}@cs.jhu.edu

Abstract. This paper describes a general-purpose system for computing registered stereoscopic video overlays of pre-operative imagery during minimally invasive surgery. There are three key elements to our approach. The first element is a real-time computer vision system that operates on stereoscopic video acquired during minimally invasive surgery to extract geometric information. We present two variations on this system: a dense stereo algorithm and a sparse point-based method. The second element is an efficient deformable surface-to-surface ICP registration. The final element is a novel display system that has been customized to operate well with stereo vision. By combining these elements, we show that we are able to perform video to volume registration and display in real time. This in turn facilitates rendering of annotations and visualization of sub-surface information on structures within the surgical field. Experimental results are shown on video sequences recorded during animal and human surgeries.

1 Introduction

Minimally invasive surgery (MIS) is a technique whereby instruments are inserted into the body via small incisions (or in some cases natural orifices), and surgery is carried out under video guidance. While advantageous to the patient, MIS presents numerous challenges for the surgeon due to the restricted field of view presented by the endoscope, the tool motion constraints imposed by the insertion point, and the loss of haptic feedback.

One means of overcoming some of these limitations is to present the surgeon with addition visual information. This paper describes a system that provides the surgeon with a three-dimensional information overlay registered to pre-operative or intra-operative volumetric data. The novelty of the system lies in its use of real-time stereo video data, online deformable registration, and rendering without recourse to an external tracking system. We have implemented a version of the system for augmenting the surgical view during laparoscopic kidney procedures.

Previous work on image overlay using stereo has largely focused on rigid structures and non-real-time visualization. In particular, [1] presents a system that makes use of stereo vision to perform image overlay of MRI images of the head. More recently, [2] briefly describes an attempt to use stereo area-matching on da Vinci images, but they largely conclude that area-matching does not work well on these images. Kanbara et al.

[3] demonstrated a video overlay using traditional image navigation techniques (which rely on an external tracking system) and rigid anatomy. Stoyanov et al. [4] described a traditional single-frame region matching stereo system and validated it against CT, and in [5] a real-time motion estimation system for discrete points was presented.

2 Methods

Briefly, our implemented system provides three general functions: 1) extraction of 3D information from stereo video data; 2) registration of video data to preoperative images; and 3) rendering and information display. We have implemented two methods for computing depth information and performing registration: a dense stereo matching algorithm, and a local point-based tracking algorithm. In all that follows, we assume that the endoscope has been calibrated to determine the corresponding 2D projection parameters [6].

2.1 Video to CT Registration

Extracting Surfaces From Stereo Most real-time stereo algorithms make use of area-matching techniques [8]. However, given the challenges of endoscopic video imagery, these techniques are often ineffective due to image noise, specular highlights and lack of texture. In recent years, global optimization methods have been developed to improve the accuracy of stereo. Original work focused on scanline dynamic programming [9]; hierarchical methods [10], multiple smoothness constraints [11], and graph cuts [12]. Our approach was to develop a highly optimized dynamic programming stereo algorithm which provides a desirable tradeoff between reliability and speed.

Fig. 1. Upper left: Standard image from stereo literature [7]; Upper right: Disparity map from upper left image (DP stereo, with sub-pixel disparity estimation and left-right check); Lower left: Image of a kidney surface; Lower right: Depth map of kidney image (DP stereo).

In the following discussion, we assume fully rectified color stereo images $L(u,v)$ (the left image) and $R(u,v)$ (the right image). In the top-left cornered image coordinate systems, for any pixel location (u,v) in the left camera, we define the disparity value for a corresponding point (u,v') in the right image as $D(u,v) = v - v'$. Given a known camera calibration, it is well known how to convert a dense disparity representation into a set of 3D points [8].

Our dynamic programming stereo algorithm optimizes the following objective function:

$$C_{tr}(u,v,d,d') = \frac{1}{2}\left(C(u-1,v,d') + C(u,v-1,d')\right) + r(d,d') + e(u,v,d) \qquad (1)$$

where d and d' are disparity values of the current pixel and the upper left neighbor, r is the cost of change in disparity between neighbors, e is an image match cost of two color pixels, and C is the cost of disparities which is initialized to 0 outside the image boundaries. The image matching function is the sum of absolute differences (SAD) of a pair of color pixels. Although the absolute difference between the two pixels proved to be adequate for most cases we observed that in certain challenging lighting conditions integrating the difference over a small region of pixels may improve matching performance. We impose a cost limit of 25 gray values which has been experimentally proven to improve matching performance [7]. The smoothness term is a linear function of the disparity values. We note that (1) is an approximation in the following sense. At a given location (u,v), it is possible that the left and upper neighbors could have differing disparities. Thus, in principle the regularization function should include separate terms for both neighbors, and the minimization in (1) should operate on two independent disparity values. However, this leads to a quadratic (in disparity) complexity. In practice, the depth resolution of stereo is far less than the lateral (pixel) resolution. As a result, in most cases there are large patches of consistent disparity. Thus, the approximation of constant local disparity is quite good, and is well worth the computational savings.

Once (1) is computed for the entire image, the disparity map satisfying both the horizontal and vertical smoothness criteria can be read out recursively from the memoization buffer M as

$$\begin{aligned} D(u_{max}, v_{max}) &= \min_d M(u,v,d) \\ D(u,v) &= \frac{1}{2}\Big[M\big(u+1,v,D(u+1,v)\big) + M\big(u,v+1,D(u,v+1)\big)\Big] \end{aligned} \tag{2}$$

In order to improve performance, we make two additional modifications. First, we choose a reduced scale (typically a factor of two to four) to perform computations. Each factor of two reduction improves performance by a factor of eight. Second, rather than searching over the complete disparity range $\mathscr{D}_s$ in every image, we only search over a small bracket of disparities about the previous stereo pair in the image sequence. With camera motion there are areas of the image (primarily at discontinuities) that occasionally violate this assumption of small change. In practice these areas converge to the correct answer within a small number of frames. As an optional feature, the algorithm is capable of computing sub-pixel disparity estimates by fitting a parabola on the costs associated to the neighbors of the winning discrete disparity. The location of the apex of resulting parabola determines the estimate of the sub-pixel disparity value.

To reduce the effects of illumination, the video data was preprocessed by first globally adjusting the brightness and color values of the left video channel to the values measured on the right channel, and then applying a a Laplacian high-boost filter in advance to increase the fine detail contrast.

Dense Registration Methods Given a 3D point cloud from stereo, a CT surface segmentation, and a good starting point (typically available based on prior knowledge of the procedure), our first goal is to compute a rigid registration (R_t, T_t) of images taken at time t to a and preoperative surface given a previous estimate (R_{t-1}, T_{t-1}).

For this purpose, we use a modified version of the classical ICP algorithm [13] applied to the depth map computed from the stereo endoscopic video stream as one

point cloud (P_{stereo}) and the 3D model of the anatomy placed in the FOV as the other point cloud (P_{model}). While P_{stereo} is a surface mesh that contains only the visible 3D details of the anatomy, P_{model} contains all the visible and occluded anatomical details. We assume that P_{stereo} is a small subset of the surface points of P_{model}, thus before finding the point correspondence the algorithm renders the z-buffer of the pre-operative model P_{model} using R_{t-1}, T_{t-1} and extracts those points that are visible by the virtual camera ($P_{modelsurface}$). The resulting $P_{modelsurface}$ point cloud is a surface mesh similar to P_{stereo}, thus finding the point correspondence with P_{stereo} is now possible. The implemented method for finding the point matches is accelerated by using a $k-d$ tree [14]. For finding the rigid transformation we used the closed-form solution technique employing SVD [15].

After finding the estimated rigid transformation using ICP, a deformable surface registration is computed. For these purposes, a set of points are defined below the surface in the CT volume, and a spring-mass system is defined as reported in [16]. For efficiency, the current implementation computes just the forces between the reconstructed surface and the CT surface. Given the point correspondence computed by the rigid transformation, we can easily compute the strain ($F(v)$) between the corresponding surface points as

$$F(v) = \gamma(P_{stereo}(CP(v)) - P_{modelsurface}(v)) \tag{3}$$

where the parameter $\gamma \in [0,1]$ determines the strength of deformation. In our results we used $\gamma = 1/3$.

In an ideal case, the strain vectors could be applied to deform the model directly. However ICP is a rigid registration algorithm thus the point correspondence between the model and the deformed surface will always be somewhat incorrect. In order to overcome this difficulty the algorithm filters the strain field (F_{filt}) before applying deformation. The filtering is done with a Gaussian kernel on the neighboring strain vectors. The neighborhood is defined in 2D on the visible surface mesh of the model. Finally the force field is applied on the model surface to yield the deformed surface ($P_{defsurface}$):

$$P_{defsurface}(v,t) = (1-\lambda)(P_{modelsurface}(v) + F_{filt}(v)) + \lambda P_{defsurface}(v,t-1) \tag{4}$$

where $\lambda \in [0,1]$ allows adjusting the temporal consistency of deformation.

Sparse Registration Methods There are many cases where surface geometry alone is inadequate for a unique, stable registration. For such cases, we have also included a stronger point-based registration method. To use this method, we first assume that the model has been brought into registration with the video, either using a dense registration or by other manual means. Once a registration is known, a set of image feature locations, $p_1, p_2, \ldots p_n$ in one image are chosen. A disparity map is calculated as described above. With this, the corresponding points in the second image are known, and the 3D locations of those points in CT coordinates are given by the registration. Thus, a direct 3D to 3D point registration can be performed using [15]. To maintain the registration, a simple brute-force template tracking algorithm has been implemented to recompute the feature points in each image. In every frame of the video, the new feature locations are used to recompute the reconstructed 3D points, and the model is re-registered.

2.2 Rendering and Display

One of the major challenges is to perform the video processing, registration, and stereoscopic rendering of the 3D overlay in real time. In order to eliminate the redundancies in the displaying and registration pipelines, a special-purpose 3D rendering engine has been developed. The current system does not use any graphics hardware acceleration for 3D rendering in order to make shared memory buffers easily accessible by the registration algorithm. This rendering engine incorporates all of the functionality of a typical graphics pipeline, including a a full geometrical transformation engine with Z-buffering, several lighting models, and various forms of transparent display. The graphics pipeline supports fast stereo rendering with no redundancy in the lighting and transformation phases, and shared texture and model memories. By sharing the same memory layout for representing the 3D geometry and having direct access to the intermediate processing steps, we can easily extract the list of visible triangles and the Z-buffer from the 3D rendering pipeline and reuse them during the dense 3D to 3D registration. The gains in memory efficiency and computational complexity are significant. The final system can render 5 million stereo triangles per second with Texture + Lighting + Transparency on a Dual Pentium 4 3.2 GHz.

During development, we asked the help of a Urologic surgeon to design the visual appearance of the 3D models so that they are visible but not obtrusive. Moreover the surgeon helped us to build other 3D models that provide additional intra-operative visual guidance for dissecting the tumor. In particular, we also display the kidney collecting system to help the surgeon understand the underlying anatomy relative to the video view. Figure 4 shows the the final display used for partial nephrectomy.

3 Results

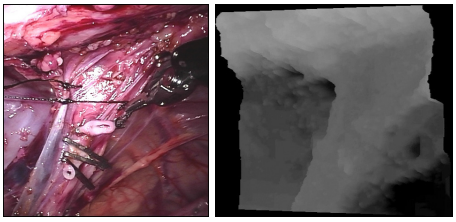


Fig. 2. Left: input image for stereo. Right: 3D mesh rendered with depth shading.

Here, we present details of the performance of the previously described stereo and registration methods. We have also included supplementary material demonstrating the system operating on representative video sequences. In terms of performance, the speed of the stereo and display engines is more than adequate for the models used in our experiments. In the case of the former, we can compute dense stereo to 1/4 pixel resolution on 240x320 images (one half VGA resolution) at over 10 frames/second. With respect to the latter, we typically have no more than 30,000 triangles on our 3D scenes which can be rendered in stereo by the engine over 100 frames per second. Both of these operate in parallel on separate CPU cores. Feature tracking is about as fast as dense stereo (10 frames/second) because large template size is required for the high accuracy feature tracking on endoscopic video data. All performance numbers include the required pre-processing computations.

3.1 Stereo and Registration on In-Vivo Data Animal Data

Dynamic Programming Stereo The dynamic programming method demonstrates very stable 3D reconstruction on an intra-operative sequence (Figure 2, left). The high depth

resolution and the fine details demonstrate that the algorithm had no difficulties dealing with the discontinuities of the anatomical surface (Figure 2 right). The only cases where significant discontinuities may occur are the areas of the surgical tools in the field of view.

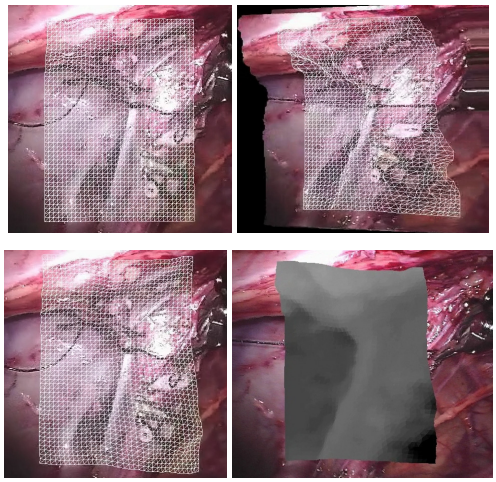


Fig. 3. Top row, rigid registration of the anatomical surface model. Bottom row, deformable registration of the anatomical surface model. On the left, the deformed wire-frame model and on the right, the deformed surface model rendered with depth shading.

Registration Results Since we did not have pre-operative 3D model of the anatomy corresponding to this surgery recording, we selected a video frame where the anatomy was not covered by any surgical tools and created a 3D surface model from the reconstructed surface mesh. We then used this model for rigid and deformable registration. As expected, the rigid registration gave perfect match for the video frame from which the model was created. For the rest of the video, frames the rigid registration provided a good approximation of the motion of the corresponding anatomical feature. ICP was configured to stop at 2 mm accuracy and process no more than 5 iterations. Enabling deformable registration improved the surface matching significantly. The deformed surface behaves like a latex surface: stretching, shrinking and sticking to the reconstructed surface (see Figure 3). For rigid registration the average error measured by ICP was below 2 mm per vertex in the video segment where the surgical tool was out of the work area (successful registration in 1 iteration). The deformable registration reduced the average registration error below 0.5 mm for most of the same video segment.

Evaluation on In-Vivo Patient Data We have applied our methods to two different interventions. Both were post-process after the surgery itself. In the first case, video data was recorded during a laparoscopic partial nephrectomy carried out using a surgical grade stereoscopic endoscope (Scholly America, West Boylston, MA). A segment of the video was chosen where the kidney surface had been exposed prior to surgical excision of the tumor. The corresponding CT image for this patient was segmented manually by a surgeon producing 3D models for the kidney surface, the tumor, and the collecting system in VTK file format.

Figure 4 shows the final display used for partial nephrectomy. The ring model that represents the cutting margins on the kidney surface around the tumor as well as the colors and the transparency levels were verified by the surgeon who performed the

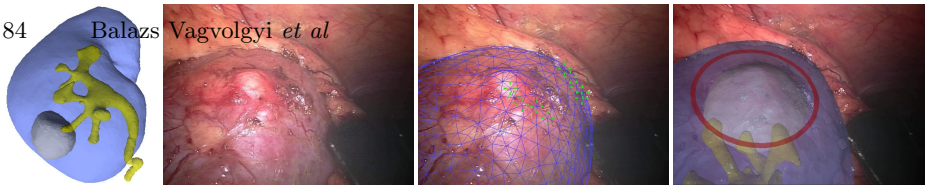


Fig. 4. Laparoscopic partial nephrectomy of a tumor (sequence 1, left to right): segmented CT model; source image (left channel); after manual registration and feature point selection; automatic registration and augmented reality overlay of the safety margin of dissection (red ring).

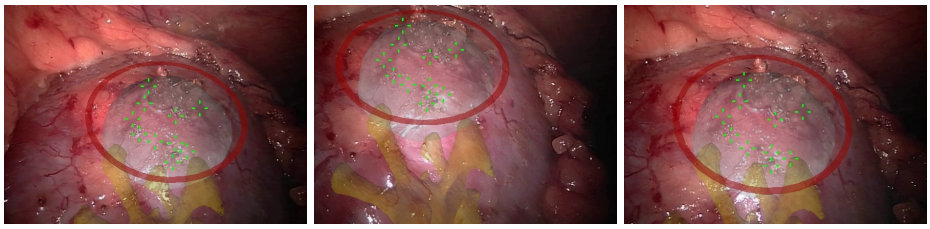


Fig. 5. Laparoscopic partial nephrectomy of a tumor (Sequence 2): automatic registration with augmented reality overlay of the safety margin of dissection (frames 154, 252, and 430).

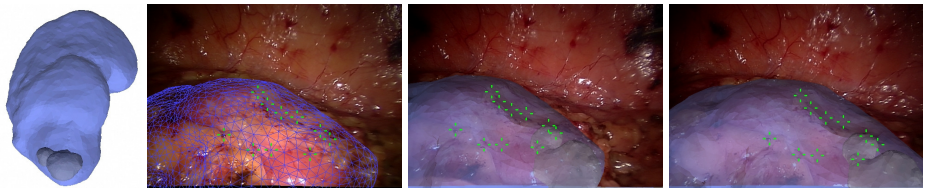


Fig. 6. Robot-assisted laparoscopic partial nephrectomy of large lower pole kidney stones (from left to right): segmented CT model; after manual registration and feature point selection; automatic registration on (frame 109); automatic registration on (frame 407).

surgery. We experimented with automated full-surface registration and manual registration followed by feature point selection and tracking. Due to the limited amount of kidney surface appearing in the video, we found that manual registration followed by "pinning" with surface feature points had superior stability as well as providing better overall performance. Figure 5 show several examples from a second sequence taken from the same case.

Our second clinical case was the surgical removal of a large kidney stone. The data was again recorded with a surgical grade stereoscopic endoscope, this time in the context of a robotic surgery carried out with the da Vinci system (Intuitive Surgical, Sunnyvale, CA). In this case, we processed approximately 1 minute of video. The CT segmentation employed did not contain the collecting system, but did contain both the stone and the kidney surface. This segmentation was also performed manually. Figure 6 shows the resulting display at three points through the video (at the video frame which provided the baseline for manual registration, and two other video frames). The associated material for this paper contains the entire video sequence. As before, we tested both the pure surface-based registration and the registration using feature points, and found the latter to be much more stable.

4 Conclusion

We have presented a system for performing real-time deformable registration and display on solid organ surfaces observed with a stereo video endoscope. The system produces good results even under the challenging conditions found in intra-operative video. We have presented results of the stereo processing and registration system on real video data, and we have evaluated the displays on two human cases.

Although a promising start, it is clear that there are several immediate avenues for further improvement of the system. First, we intend to combine the surface and local feature tracking registration, and to automate the selection of points for the latter. Second, we are working to parallelize the algorithm to improve the speed of both stereo processing and registration. Finally, we are planning to perform a formal system validation in an animal model within the next few months.

Acknowledgements This material is based upon work supported by TATRC under grant W81XWH-06-1-0195 and the National Science Foundation under Grant No. EEC-9731478. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors would like to thank Carol Reiley for the video recordings.

References

[1] Betting, F., Feldmar, J., Ayache, N., Devernay, F.: A new framework for fusing stereo images with volumetric medical images. In: CVRMed. (1995) 30–39
[2] Mourgues, F., Devernay, F., Coste-Maniere, E.: 3d reconstruction of the operating field for image overlay in 3d-endoscopic surgery. isar **00** (2001) 191
[3] M. Kanbara et al.: A stereoscopic video see-through augmented reality system based on real-time vision-based registration. In: Proc. Virt. Reality. (2000) 255–262
[4] Stoyanov, D., Darzi, A., Yang, G.Z.: Dense 3d depth recovery for soft tissue deformation during robotically assisted laparoscopic surgery. In: MICCAI (2). (2004) 41–48
[5] 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: MICCAI (2). (2005) 139–146
[6] Zhang, Z.: A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence **22**(11) (2000) 1330–1334
[7] Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision **47**(1) (2002) 7–42
[8] Brown, M.Z., Burschka, D., Hager, G.D.: Advances in Computational Stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence **25**(8) (2003) 993–1008
[9] Cox, I.J., Hingorani, S.L., Rao, S., Maggs, B.M.: A maximum likelihood stereo algorithm. Computer Vision and Image Understanding **63**(3) (1996) 542–567
[10] van Meerbergen, G., Vergauwen, M., Pollefeys, M., Van Gool, L.: A hierarchical symmetric stereo algorithm using dynamic programming. IJCV **47**(1-3) (2002) 275–285
[11] Hirschmüller, H.: Stereo vision in structured environments by consistent semi-global matching. In: Proc. CVPR. (2006) 2386–2393
[12] Kolmogorov, V., Zabih, R.: Graph cut algorithms for binocular stereo with occlusions. In: Mathematical Models in Computer Vision: The Handbook. Springer-Verlag (2005)

[13] Besl, P., McKay, N.: A method for registration of 3D shapes. PAMI **14**(2) (1992) 239–256
[14] Williams, J.P., Taylor, R.H., Wolff, L.B.: Augmented k-d techniques for accelerated registration and distance measurement of surfaces. In: Computer Aided Surgery: Computer-Integrated Surgery of the Head and Spine, Linz, Austria (1997) P01–21
[15] Arun, K.S., Huang, T.S., Blostein, S.D.: Least-squares fitting of two 3-D point sets. IEEE Trans. Pat. Anal. Machine Intell. **9** (1987) 698–700
[16] K. Montgomery et. al.: Spring: A general framework for collaborative real-time surgical simulation. In: Proc. MMVR. (2002)

Integrating Autostereoscopic Multi-View Lenticular Displays in Minimally Invasive Angiography

Daniel Ruijters

X-Ray Predevelopment, Philips Healthcare, Best, the Netherlands
danny.ruijters@philips.com.

Abstract. In order to gain maximal insight in the vascular morphology during minimal invasive treatment of vascular pathologies, high resolution 3D Rotational Angiography (3DRA) data is acquired and visualized peri-operatively. The 3D data provides good understanding of the vasculature, however in case of complex structures (*e.g.* liver and neuro-vasculature, and arteriovenous malformations) there is a distinct added value in showing a stereoscopic visualization to the clinician. In order to provide such a stereoscopic image, a 42" autosterescopic nine-view lenticular screen was mounted in the Operating Room (OR). To optimally assist the intervention, the stereoscopic visualization of the vasculature can be presented from the same viewing incidence as the real-time X-ray image, which is used to guide the intra-vascular devices, such as catheters and stents. The GPU-based implementation offers interactive manipulation throughout.

1 Introduction

Autostereoscopic displays allow a stereoscopic view of a 3D scene, without the use of any external aid, such as goggles. The additional depth impression that a stereoscopic image offers, allows a natural interpretation of 3D data. Principally there are two methods for conveying a stereoscopic image: time multiplexing and spatial multiplexing of two or more views. Though two views are enough to create the impression of depth (after all, we have only two eyes), offering more views has the advantage that the viewer is not restricted to a fixed sweet spot, since there is a range of positions where the viewer will be presented with a stereoscopic visualization. As a consequence, multiple viewers can look at the same stereoscopic screen, without wearing goggles. Furthermore it is possible to 'look around' an object, when moving within the stereoscopic range, which aids the depth perception. The multi-view lenticular display uses a sheet of lenses to spatially multiplex the views [1], and typically offers four to fifteen spatially sequential images.

The Graphics Processing Unit (GPU) is a powerful parallel processor on today's off-the-shelf graphics cards. It is especially capable in performing Single Instruction Multiple Data (SIMD) operations on large amounts of data. In this

article a method for rendering a 3D scene for display on lenticular screens is described, benefitting from the vast processing power of modern graphics hardware. The presented approach allows dynamical adjustment of the resolution of the rendered images, in order to guarantee a minimal frame rate. In this way interactive frame rates can be reached, enabling to fully profit from lenticular displays in a clinical environment.

2 State of the art

In 1838 Sir Charles Wheatstone developed a device, called the stereoscope, which allowed the left and the right eye to be presented with a different image, in order to create a impression of depth. The development of *auto*stereoscopic display devices, presenting stereoscopic images without the use of glasses, goggles or other viewing aids has seen an increasing interest since the 1990s [2–4].

The advancement of large high resolution LCD grids, with sufficient brightness and contrast, has brought high quality multi-view autostereoscopic lenticular displays within reach [5]. The range of viewing positions, allowing the perception of a stereoscopic image, is mainly determined by the number of views, offered by the display.

Several publications describe how the GPU can be employed to extract the data stream for the lenticular display from a 3D scene in an effective manner. Kooima *et al.* [6] and Kratz *et al.* [7] present a two-pass GPU based algorithm for a two-view head-tracked stereoscopic display. First the views for the left and the right eye are rendered, and in the subsequent pass they are interweaved. Domonkos *et al.* [8] describe a two-pass approach, dedicated for iso-surface rendering. In the first pass they perform the geometry calculations on the pixel-shader for every individual pixel, and in the second pass the shading is performed. Hübner and Pajarola [9] describe a GPU-based single-pass multi-view volume rendering, varying the direction of the casted rays depending on their location on the lenticular screen.

The previous GPU-based approaches were dedicated render methods, working on the native resolution of the lenticular LCD grid. The approach in this paper decouples the rendering resolution from the native LCD grid resolution, allowing lower resolutions, when higher frame rates are demanded.

3 The multi-view lenticular display

The multi-view lenticular display device consists a sheet of cylindrical lenses (lenticulars) placed on top of an LCD in such a way that the LCD image plane is located at the focal plane of the lenses [10]. Therefore, LCD pixels located at different positions underneath the lenticulars fill the lenses when viewed from different directions, see figure 1. Provided that these pixels are loaded with suitable stereo information, a 3D stereo effect is obtained, in which the left and right

eye see different, but matching information. The screen we used offered nine distinct angular views, but the described method is applicable to any number of views.

(a) (b) (c)

Fig. 1. (a) Stereoscopic image perception. Two different views are combined by the brain, in order to perceive depth. (b) The light of the sub-pixels under the lenticular lenses is directed into different directions [11]. (c) The cylindrical lenses depict every sub-pixel in a different view. The numbers in the sub-pixels indicate in which view they are visible.

The fact that the different LCD pixels are assigned to different views (spatial multiplex), leads to a lower resolution per view than the resolution of the LCD grid [12]. In order to multiplex the views horizontally (our eyes are placed horizontally), the lenticular cylindrical lenses should be placed vertically, parallel to the LCD pixel grid columns. However, this would reduce the resolution only over the horizontal axis, which is undesirable. Therefore, the lens cylinders are slanted at a small angle, distributing the the reduction of resolution over the horizontal and vertical axis [1]. The resulting assignment of a set of LCD pixels, is illustrated in figure 1c. Note that the red, green and blue color channels of a single pixel are depicted in different views. Every view offers a full color image, though.

4 Method

We use a two pass algorithm: First the individual views from the different foci positions are separately rendered to an orthogonal grid. In the second pass, the final output signal has to be resampled from the views to a non-orthogonal grid in the compositing phase (see figure 2). The processing power of the GPU is harvested for both passes. In order to maintain an acceptable frame rate, the resolution of the views can be changed dynamically.

The frustums that result from the different focal spots, are illustrated in figure 3. The viewing directions of the frustums are not parallel to the normal of the screen, except for the center one. Therefore the corresponding frustums are asymmetric [13]. A world coordinate (x, y, z) that is perspectively projected, using

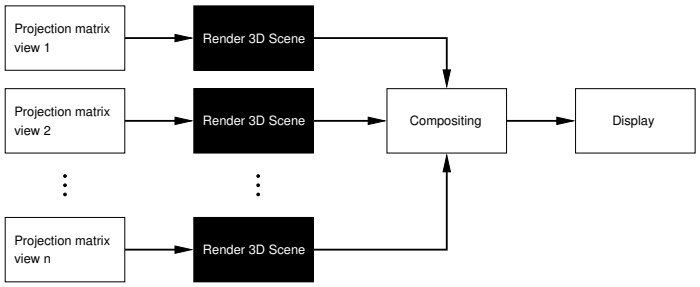
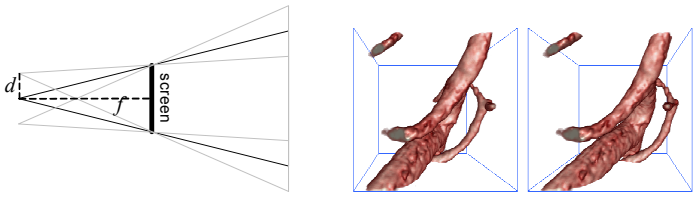


Fig. 2. The process of rendering for the lenticular display. Optionally, the rendering of the N individual views can be done in parallel.



(a) (b)

Fig. 3. (a) The frustums resulting from three different view points. (b) The same scene rendered from the extreme left and right viewpoints.

such an asymmetric frustum, leads to the following view port coordinate $v(x, y)$:

$$v(x,y) = \left(\frac{(x - n \cdot d) \cdot f}{f - z} + n \cdot d, \quad \frac{y \cdot f}{f - z} \right) \tag{1}$$

Whereby f denotes the focal distance, n the view number and d the distance between the view cameras.

After the projection matrix has been established, using equation 1, the scene has to be rendered for that particular view. All views are stored in a single texture, which is called *texture1*. In OpenGL, the views can be placed next to each other in horizontal direction, using the `glViewport` command. The location of a pixel in view n in *texture1* can be found as follows:

$$\boldsymbol{t} = \left(\frac{1}{2} + \frac{n}{N} + \frac{2p_x - 1}{2N}, \quad p_y \right) \tag{2}$$

whereby $\boldsymbol{t}$ denotes the normalized texture coordinate, $\boldsymbol{p}$ the normalized pixel coordinate within the view, and N the total number of views. The view index n is here assumed to be in the range $[-\frac{1}{2}(N-1), \frac{1}{2}(N-1)]$, as is used in figure 1c.

To composite the final image, which will be displayed on the lenticular screen, the red, green and blue component of each pixel has to be sampled from a different view. The view number stays fixed all the time for each sub-pixel,

therefore this information is pre-calculated once, and then put in a static texture map, called *texture0*.

In the compositing phase, all the pixels in the output image are parsed by a GPU program. For each normalized pixel coordinate $\boldsymbol{p}$ in the output image, *texture0* will deliver the view numbers n that have to be sampled for the red, green and blue components. The respective views are then sampled in *texture1* according equation 2 using bi-linear interpolation, delivering the appropriate pixel value. A performance characterization of volume rendering to the lenticular screen according to the presented approach is given in table 1.

Volume description	Frames per second	Views per second
3DRX foot, $256^2 \cdot 200$ voxels (25 MB)	52.4	471
CT head, $512^2 \cdot 256$ voxels (128 MB)	19.2	173
3DRA neuro-vascular (sparse), 512^3 voxels (256 MB)	21.3	192
4D cardiac CT, 3 phases of $512^2 \cdot 333$ voxels (totally 510 MB)	13.6	123

Table 1. In order to characterize the performance of GPU-accelerated volume rendering [14] and compositing on the 9 view lenticular screen, several data sets were rendered at the maximum resolution to an output window of 800^2 pixels. All measurements were obtained, using a 2.33 GHz Pentium 4 system, with 2 GB RAM memory, and a nVidia QuadroFX 3500 with 256 MB on board memory as graphics card.

5 Clinical setup

We applied the presented approach to visualize intra-operatively acquired 3D data sets on a Philips 42" lenticular screen, which was mounted in the operation room (OR). The orientation of the depicted 3D data set can follow in realtime the orientation of an X-ray C-arc system (see figure 4), which means that on the lenticular display the 3D data set is visualized from the same viewing angle, as the viewing incidence on the patient in the real-time X-ray image. This approach allows to reduce the X-ray radiation, since the physician can choose the optimal orientation to acquire X-ray images without actually radiating. Further it improves the interpretation of the live projective 2D X-ray image, which is presented on a separate display, since the 3D data on the stereoscopic screen (which is in the same orientation) gives a proper depth impression through the stereovision of the lenticular screen. The fact that the clinician is not limited to single sweet spot, makes the multi-view display particularly suitable for this environment, since the clinical intervention demands that an operator can be positioned freely in the range close to the patient.

The ability to visualize and manipulate the 3D data interactively is of great importance in the analysis and interpretation of the data. Interactivity, in this

Fig. 4. Left: the 42" lenticular screen (top right display) in the operating room. Middle: the X-ray C-arc system in a clinical intervention. Right: the degrees of rotational freedom of the C-arc geometry.

Fig. 5. Volume rendered vascular 3DRA images. From left to right: brain arteriovenous malformations, virtual stenting and aneurysm (blue), another neuro vessel tree with aneurysm, and vasculature mixed with soft-tissue data.

context, means that the frame rates of the visualization are sufficient to provide direct feedback during user manipulation (such as rotating the scene). When the visualization's frame rate is too low manipulation becomes very cumbersome. Five frames per second are often used as a required minimum frame rate.

Especially the cerebral vasculature consists of many curved vessels, see figure 5. From a single X-ray image it is impossible to interpret the in-plane curvature (the curvature in the z-direction of the image). But even looking at a 3D rendered image on a 2D plane (*i.e.* conventional monitor), it is often very difficult to estimate the in-plane curvature without rotating the vessel tree. Rotating a 3D scene is not a problem when sitting behind a desktop computer, but during a clinical intervention the performing clinician is primarily occupied with the medical procedure, and interaction with sterile computer input devices (which are available in the OR) is an additional task, demanding focus. The stereoscopic image allows to interpret the 3D shape, including the in-plane curvature, in a single glance without any additional input interaction, and therefore reduces the mental stress on the clinician during the intervention.

5.1 Follow C-arc

When the Follow C-arc mode is active, the viewing incidence of the 3DRA data set is matched in real-time with the current rotation and angulation of the C-arc geometry (see figure 4). This allows the clinician to use the view on the 3DRA data set to predict an optimal working position for the C-arc without actually using X-ray radiation.

In order to perform this, the transformation of the coordinate space of the 3DRA data to the coordinate space of the C-arc X-ray detector has to be established. Let the origin of the coordinate system of the 3DRA data be positioned at the center of the data set, and let the x-axis correspond to the short side of the table, the y-axis to the long side of the table, and the z-axis point from the floor to the ceiling. The X-ray C-arc system can rotate over three axis (see figure 4b). The rotation of the detector coordinate system, with respect to the 3DRA data can be expressed as:

$$M = R_x \cdot R_y \cdot R_z =$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix} \cdot \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix} \cdot \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3)$$

Note that the order of the matrix multiplications is given by the mechanics of the C-arc system. The C-arc system's iso-center serves as origin for the rotation matrices. The rotation of 3DRA volume to the detector coordinate system corresponds to the inverse of matrix M, which is equal to its transposed, since rotation matrices are orthogonal.

The counterpart of the follow C-arc function is called 3D APC (Automatic Positioning Control), which allows the C-arc to be moved to a viewing incidence corresponding to the orientation of the 3D rendering of the 3DRA data set. Desired working positions can be planned and stored before or during the intervention and later recalled during the intervention when needed.

6 Conclusions

In this article an approach was described, bringing stereoscopic visualizations in the operating room, in order to support the clinician during minimal invasive treatment of vascular pathologies using X-ray angiography. We used a 42" nine-view lenticular screen to offer an image that could be easily interpreted by the clinicians. The orientation of the 3DRA data on the lenticular display followed the viewing incidence of the X-ray C-arm in real-time, in order to provide maximum insight in the anatomy that is depicted on a corresponding X-ray image.

Due to the GPU-acceleration, together with the adaptive adjustment of the intermediate view resolution, interactive frame rates can be reached, which allows intuitive manipulation of the rendered scene. Since both the 3D rendering and

the compositing take place on the graphics hardware, the requirements for the other components of the PC system are rather modest. Thus the realization of the proposed high performance system can be very cost effective.

First feedback from the clinicians indicates that there certainly is a perspective for clinical added value. Orthopedic surgery is suggested as another application area that could benefit from the multi-view stereoscopic display. For better integrated usage, the display should be mounted on the same ceiling suspension with the other (2D) displays. Also the integration of the live fluoroscopy image in the 3D scene would be highly appreciated. The fact that the clinicians do not need to wear any additional glasses, and are not limited to a sweet spot, as well as the fact that large data sets can be manipulated interactively, make this method very suitable for a clinical interventional environment.

References

1. van Berkel, C., Parker, D.W., Franklin, A.R.: Multiview 3D LCD. In: Proc. SPIE - Stereoscopic Displays and Virtual Reality Systems III. Volume 2653. (April 1996) 32–39
2. Halle, M.: Autostereoscopic displays and computer graphics. ACM Computer Graphics **31**(2) (May 1997) 58–62
3. Dodgson, N.A.: Autostereoscopic 3D displays. Computer **38**(8) (2005) 31–36
4. Onural, L., Sikora, T., Ostermann, J., Smolic, A., Civanlar, M.R., Watson, J.: An assessment of 3DTV technologies. In: Proc. NAB Broadcast Engineering. (2006) 456–467
5. Vetro, A., Matusik, W., Pfister, H., Xin, J.: Coding approaches for end-to-end 3D TV systems. In: Picture Coding Symposium (PCS). (December 2004)
6. Kooima, R.L., Peterka, T., Girado, J.I., Ge, J., Sandin, D.J., DeFanti, T.A.: A GPU sub-pixel algorithm for autostereoscopic virtual reality. In: IEEE Virtual Reality Conference. (Mar 2007) 131–137
7. Kratz, A., Hadwiger, M., Fuhrmann, A., Splechtna, R., Bühler, K.: GPU-based high-quality volume rendering for virtual environments. In: Proc. AMI-ARCS. (Sep 2006)
8. Domonkos, B., Egri, A., Fóris, T., Juhász, T., Szirmay-Kalos, L.: Isosurface raycasting for autostereoscopic displays. In: Proc. WSCG. (January 2007) 31–38
9. Hübner, T., Pajarola, R.: Single-pass multi-view volume rendering. In: Proc. IADIS International Computer Graphics and Visualization. (July 2007)
10. van Berkel, C.: Image preparation for 3D-LCD. In: Proc. SPIE - Stereoscopic Displays and Virtual Reality Systems VI. Volume 3639. (May 1999) 84–91
11. Braspenning, R., Brouwer, E., de Haan, G.: Visual quality assessment of lenticular based 3D displays. In: Proc. 13th European Signal Processing Conference (EUSIPCO). (September 2005)
12. Dodgson, N.A.: Autostereo displays: 3D without glasses. In: EID: Electronic Information Displays. (1997)
13. Maupu, D., Horn, M.H.V., Weeks, S., Bullit, E.: 3D stereo interactive medical visualization. IEEE Computer Graphics and Applications **25**(5) (Sept.-Oct. 2005) 67–71
14. Ruijters, D., Vilanova, A.: Optimizing GPU volume rendering. Journal of WSCG **14**(1-3) (January 2006) 9–16

In-Situ Visualization of Medical Images Using Holographic Optics

John M. Galeotti[1], Mel Siegel[1], and George Stetten[1,2]

[1] Robotics Institute, Carnegie Mellon University, USA,
galeotti+miccai@cs.cmu.edu,
[2] Department of BioEngineering, University of Pittsburgh, USA.

Abstract. We present the first autostereoscopic visualization system to project a real-time (live) medical data slice in-situ by use of a holographic optical element. Our system can project an essentially (1 mm in-slice accuracy) viewpoint-independent real-time virtual image into its actual anatomic location, enabling natural hand-eye coordination to guide invasive procedures. Our system does not require tracking or a head-mounted device, and it can project a 104x112 mm virtual image from a much smaller 19.3x15.5 mm source image. The system's optics are non-axial, arranged so that the source image does not block the direct view of the patient and allows sufficient room for long tools to operate on the patient. We are currently adapting this system for use with ultrasound for guiding liver biopsy and amniocentesis.

1 Introduction

In the current practice of medicine, images are routinely acquired by ultrasound, computerized tomography (CT), magnetic resonance imaging (MRI) and other modalities. These images are viewed on a film or screen, rather than by looking directly into the patient. This separation between image display and the patient workspace requires a clinician using the images for real-time guidance to mentally integrate two disparate frames of reference. The difficultly in performing such mental integration is problematic for performing invasive procedures, where direct physical interaction with the region being imaged is required. Examples of such procedures include vascular access, biopsy, amniocentesis, and minimally invasive (keyhole) surgery.

A number of researchers have worked to develop more natural ways to merge images with the perceptual real world, thereby removing the clinician's need to shift their gaze between the patient and the image [1–7]. These techniques fall into the broad category of *augmented reality* (AR), which enhances what is predominantly a real scene with virtual objects. For example, in the case of guiding needle biopsy, both the patient and needle are real. The "virtual" addition to the scene may consist of a real-time ultrasound image being used to guide the biopsy procedure or an MR or CT image acquired previously. AR seeks to project such images *in situ* (in the location from which they were scanned).

The most common method for displaying an in-situ object in current commercial and research AR systems is to present a separate viewpoint-dependent rendering to each eye, such that stereoscopic vision may determine depth [1–3]. Doing so requires both real-time computation to generate an appropriate rendering for each eye, as well as some sort of tracking to determine each eye's 3D position. Many such systems make use of either a head-mounted display (HMD) or special, e.g. polarized, glasses. All of these systems have a number of difficulties, some of which are technical hurdles but others of which are intrinsic, such as conflicting depth cues (e.g. accommodation and convergence) [8].

Optical systems that generate in-situ images without tracking or a head-mounted display offer a solution to many of these difficulties. The perceived 3D location of each in-situ point in these systems is essentially independent of viewpoint, allowing natural depth perception of the 3D scene. Accordingly, these systems are *autostereoscopic* [9]. Such systems may either project a *holographic image* or they may project a true optical virtual image, in the latter case either by means of a semi-transparent mirror or, as presented here, by means of a *holographic optical element* (HOE). In contrast to a holographic image, an HOE is a hologram of an optical system that produces the desired image projection (see Section 2).

1.1 Holograms for AR

Holographic images are naturally autostereoscopic and well-suited for interventional guidance. Significantly, they can be merged with a direct view of the patient, allowing the clinician to see both the in-situ holographic image as well as where they are puncturing the skin. Such merger of direct vision with in-situ visualization is achievable by use of a narrow band hologram, which allows the bulk of the visible spectrum to pass directly through unmodified.

Significant research has been done in the area of real-time projection of 3D holographic images. A recent break-through has been the development of a re-writeable holographic recording medium by Tay, et al. [10]. Their new photorefractive polymer composite changes its refractive index in response to the recording laser via an electrical rather than chemical process, allowing the creation of a hologram that can be repeatedly erased and re-recorded. Their process is currently slow, requiring approximately 2 minutes to record an image, but future advances could potentially allow a sufficiently powerful laser to record holograms at video frame rates. Even if such a feat were to be achieved, however, there is a remaining group of problems for using holographic images in augmented reality medical applications.

1.2 Difficulties Utilizing Holographic Images for AR

The preferred method of creating a holographic image from an acquired medical image is to compute the appropriate phase patterns for each location on the hologram (hogel) directly from a computer model, and then individually write each hogel (utilizing a spatial light modulator) as described in [10]. There are

two potential hurdles to this approach. First, the resulting tessellation of the hologram into small hogels (each of which records only one 2D perspective of the 3D object to be displayed) quantizes the desired parallax effects and stereoscopic depth perception. Second, it can be time consuming both to compute the plethora of perspectives and to individually (i.e., sequentially) write the large number of hogels to the hologram media.

The second challenge can sometimes be mitigated by only computing and recording perspectives across the horizontal axis [10]. Even when such a hologram is properly aligned with the viewer, vertical changes in viewpoint will lead to different perceived locations for the hologram image along the vertical axis. This may be acceptable for "Princess Leia" floating in space, but could be disastrous if used to guide an invasive procedure. Invasive procedure guidance requires accurate localization along all three axes, as clinicians may well tilt their heads or move them vertically as they bend over a patient in the normal course of performing an intervention.

Taking these hurdles together, it becomes apparent that for the foreseeable future *hologram recording is unamenable to real-time updating* to match ongoing operational scans, e.g. from ultrasound. The time required to update a 2D hogel array will also preclude holographic video of pre-acquired periodic time sequences of images, e.g. cardiac gated volumetric CT. Furthermore, even if these hurdles are overcome, future lasers suitable for real-time hologram writing will likely be large and high-powered, making it difficult to construct a safe hand-held system that does not block the operator's vision or place physical equipment in the way of surgical tools.

1.3 Other Autostereoscopic In-Situ Visualization Methods

An alternative approach is to project a true optical virtual image, which occurs as the result of the *apparent* in-focus convergence of light rays, such as a reflection in a mirror or the magnified image produced by a magnifying glass. *Like holographic images, true optical virtual images are naturally autostereoscopic.* This approach is capable of autostereoscopic real-time operation using current technology. Real-time tomographic reflection (RTTR), such as is employed by Stetten's sonic flashlight, is one such AR technique [5–7]. As shown in Fig. 1, RTTR utilizes a half-silvered (semi-transparent) mirror to project an in-situ virtual image from a real-time image source (such as an LCD or OLED display). By rigidly connecting the autostereoscopic visualization device to the scanning device, the need for cumbersome tracking equipment is altogether avoided.

Until recently, RTTR was the only technique available for projecting in-situ autostereoscopic virtual images in real time. Unlike holography, however, RTTR is typically restricted to displaying a 2D manifold (albeit corectly located and perceived in 3D space) of the same size, shape, and mirrored position as the display source. Accordingly, it can be difficult or impossible to construct a hand-held RTTR system that is not unwieldy but is still capable of guiding deeper procedures such as liver biopsy or amniocentesis. A hand-held form factor is especially desirable for use with portable scanning technologies, such as ultrasound.

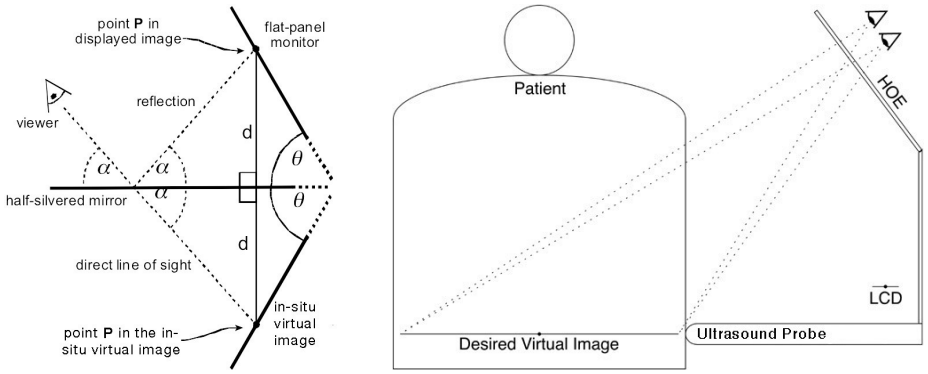


Fig. 1. Left: RTTR Configuration: A half-silvered mirror bisects the angle between the in-situ virtual image (coincident with the scanned data, e.g. ultrasound slice, within the patient) and the flat-panel monitor. Point P in the virtual image and its corresponding location on the monitor are equidistant from the mirror along a line perpendicular to the mirror (distance = d). Because the angle of incidence equals the angle of reflection (angle = α) the viewer (shown as an eye) sees each point in the reflection precisely at its corresponding physical 3D location, independent of viewer location. **Right:** An HOE can be used to project an autostereoscopic virtual image whose size, shape, and position need not be identical to that of the display source.

2 Holographic Optical Elements

There is a need for an AR method that is not only *presently* capable of real-time, in-situ autostereoscopic visualization of "large" objects, but is also capable of doing so from a hand-held device without using head tracking, blocking the operator's vision, or placing physical equipment in the way of surgical tools. Holography is not yet capable of real-time operation, and it will not be implementable in a hand-held form factor for the foreseeable future. RTTR has already been embodied in a real-time hand-held device (the sonic flashlight), but is capable of visualizing only 2D manifolds of limited size without resorting to an unwieldy form factor and/or blocking physical access to the patient by placing a mirror in the way of surgical tools. Here we present an alternative approach, originally proposed (but unimplemented) in [11].

An HOE can be used in lieu of RTTR's semi-transparent mirror, potentially allowing the creation of a hand-held device with the above capabilities. This is possible because a single narrow-band HOE can appear transparent, and yet still combine the magnifying capabilities of a lens with the "repositioning" capabilities of a diffraction grating to project a large in-situ true optical virtual image. The content of the virtual image is generated from a semi-arbitrarily positioned small real-time image source, such as an LCD back-light by a laser. Thus, an HOE is capable of enabling a device such as depicted on the right in Fig. 1.

An HOE can be conceptualized as a permanently recorded hologram, whose purpose is not to project a fixed 3D image, but rather to act as an optical element, such as a lens or a prism. In the present case, the purpose of the HOE is to project an autostereoscopic virtual image containing the data displayed on a separate real-time image source. Real-time operation is achievable because the image source, not the HOE, is updated in real time to visualize each new image "frame." The visualization of larger objects by way of an uncumbersome hand-held device is possible due to the HOE's diffraction-based operation and lack of reliance on a large, high-quality laser. Finally, an unobstructed view of the patient is possible through the HOE if the HOE is sufficiently narrow-band.

As with RTTR, the use of a 2D image source unfortunately restricts an HOE-based system such that, at any given moment, it can only visualize data lying on a 2D manifold. The projected tomographic data can be correctly located and perceived in 3D space, and different manifolds of data may be examined in temporal sequence by physically moving the device (and thus the location of the in-situ visualization). However, simultaneous visualization of a 3D volume is not possible without, at minimum, the use of multiple optical elements and a high-speed image source to automatically project a temporal and spatial series of virtual image "slices" at known locations and in rapid succession. Even so, present technology is either at or near the capability required to project a stack of virtual images in real time, and so HOE design (as opposed to holography improvements) may well usher in the era of 3D real-time autostereoscopic visualization.

Even without volumetric 3D displays, HOE-based autostereoscopic visualization systems may soon be capable of guiding deeper procedures such as liver biopsy or amniocentesis by projecting real-time 2D ultrasound in situ, and future HOE-based 3D visualization systems may be capable of visualizing arbitrary volumetric data in real time, such as may be acquired by 3D ultrasound, OCT, or even preoperative MRI or CT registered in real-time to ultrasound, etc. Not only would these new visualization capabilities have the potential to improve the quality of care for existing procedures, but new interventional procedures may also be enabled, such as out-of-plane needle insertion to reach problematically located targets.

3 Method

There are several types of HOE, each with a different fabrication method. Regrettably, no HOE is capable of projecting a perfect virtual image; there will always be some degree of optical aberration that results in blur. Optical aberrations are inherent to any non-trivial optical component[3], whether refractive such as a lens or diffractive such as an HOE. Unfortunately, blur in a virtual image degrades its autostereoscopic quality, because a projected point that is classically blurred over the entire aperture of an HOE will be perceived as being in different

[3] A perfectly flat front-surface mirror can be aberration-free.

locations when viewed through different regions of the HOE. Accordingly, one of the major goals in designing an HOE-based in-situ visualization system is the minimization of blur, not only to produce images that appear sharp, but also to reduce the undesired dependence of apparent target location upon viewpoint.

In general, an HOE-based system is designed using optical simulation and optimization software (in particular, we have made use of the commercial Zemax EE optical engineering program). The HOE's phase function, which fully describes, e.g., the refractive index across the HOE's surface, is usually jointly optimized with the remainder of the device's optical components, including their physical layout. Such optimization is guided by an often-complex merit function, designed to favor physically realizable designs while minimizing optical aberrations. Due to their complexity, merit functions are typically subject to many local minima. Accordingly, such optical design is both an art and a science, requiring a carefully crafted merit function and tedious coaxing of the system through the solution space by means of many carefully chosen intermediate designs (and corresponding merit functions).

A complete HOE-based visualization system requires, at minimum, an HOE and a display source, such as an LCD. Each pixel of the display source must emanate diffused light at the HOE's operating wavelength, which may be accomplished by use of an expanded laser beam for illumination paired with an optical diffuser. In order to help correct for the HOE's optical aberrations, it may be necessary to place additional corrective optics between the image source and the HOE. To avoid either blocking the operator's vision or placing physical equipment in the way of surgical tools, such corrective optics must themselves be kept out of the way. The properties of these corrective optics are jointly optimized with the rest of the optical system. The set of corrective optics may include a fiber optic faceplate located adjacent to the display source. The fiber optic faceplate can serve dual purposes, correcting field curvature aberration while simultaneously functioning as the afore-mentioned optical diffuser.

4 Results

We have designed and constructed one such HOE-based system (see Fig. 2), capable of projecting a virtual image sized 104 mm wide by 112 mm tall, 5.4 times wider and 7.2 times taller than our 19.3x15.5 mm image source. To simplify our initial design, we positioned the virtual image at 1 m from the HOE, further than likely to be used in clinical practice (reducing this distance increases optical aberrations). Our design has an unobstructed line of sight from the HOE to the virtual image, with the image source and two corrective lenses located off-axis, out of the way of the direct line of sight to the patient.

The projected virtual image is well focused, and thus objects in the virtual image do not appear to substantially change their position across the range of viewpoints. In particular, over a normal range of viewpoints, objects empirically do not appear to move by more than 1 mm within the image slice, as quantitatively measured using a digital camera on a precision positioning stage and

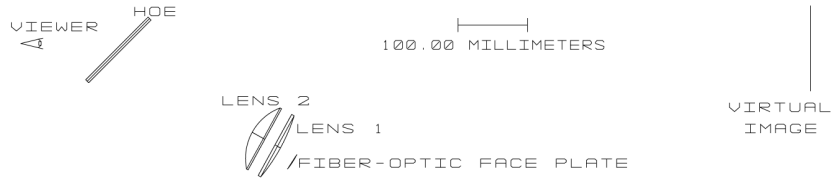


Fig. 2. The layout of our HOE-based autostereoscopic in-situ visualization system.

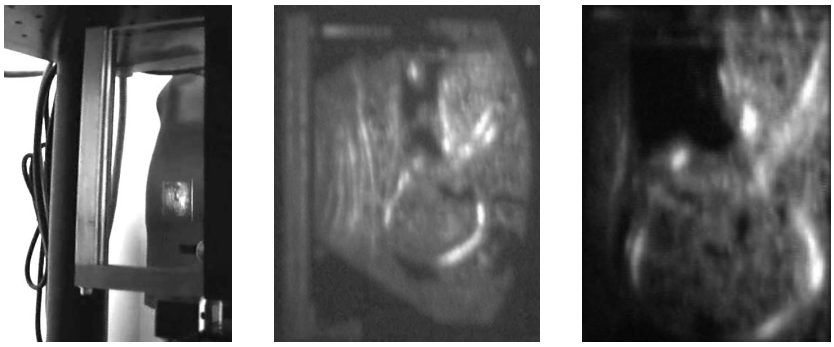


Fig. 3. Photos showing fetal ultrasound projected inside the mother. The images appear to have higher resolution in person, due primarily to the larger dynamic range and smaller pupil size of the human visual system. The left photograph shows the ultrasound image floating inside the mother, as viewed through the HOE. The other photographs are "zoomed in," showing part of the fetus, oriented facing left with the head down. Some of these are still-shots from a video we recorded showing in-situ autostereoscopic visualization of fetal movement inside the mother. For this demonstration, we played back previously recorded ultrasound. However, by simply attaching an ultrasound machine to our current device we could project real-time ultrasound data in situ.

well in agreement with our optical simulation. Actual in-situ projection of fetal ultrasound is shown in Fig. 3.

5 Conclusion

In-situ guidance of deep procedures such as liver biopsy or amniocentesis requires the real-time projection of large virtual images of 2D (e.g., ultrasound) or 3D data in situ, without blocking the operator's vision or placing physical equipment in the way of surgical tools.

We hypothesized that HOE-based autostereoscopic visualization systems are capable of meeting these requirements. Rather than seeking to project true 3D holographic images (and thus requiring real-time updates of a hologram), we use

a permanently recorded HOE to project a real-time virtual image from a readily available image source such as an LCD (the LCD, not the HOE, is updated in real time). Unlike RTTR systems, the size, shape and position of the virtual image are not tightly constrained by the size, shape, and position of the image source. In the future, volumetric 3D visualization may also be possible.

We have designed, built, and tested the first such HOE-based AR system, successfully demonstrating the chief benefits of using an HOE.

Having shown that HOE-based in-situ guidance of deep procedures is possible, we are now working on a new design with a larger virtual image and a reduced distance between the HOE and the virtual image, which are more desirable for hand-held in-situ projection of ultrasound. Our new optical design is more challenging, because the increased image size and reduced optical distance exacerbate many optical aberrations. Once our design is finished, however, our new autostereoscopic device should be well-suited for guiding deep invasive procedures.

Acknowledgments

We are thankful to Andreas Nowatzyk, Richard Rallison, and Wasatch Photonics for their ideas, suggestions, and assistance with this work. This work was funded in part by NIH R21 EB007721 and by NSF 0308096.

References

1. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., MacIntyre, B.: Recent advances in augmented reality. IEEE Computer Graphics and Applications **21**(6) (2001) 34–47
2. Fuchs, H., State, A., Pisano, E., Garret, W., Hirota, G., Livingston, M., Whitton, M., Pizer, S.: Towards performing ultrasound-guided needle biopsies from within a head-mounted display. In: Visualization in Biomedical Computing, Hamburg, Germany (1996) 591–600
3. State, A., Ackerman, J., Hirota, G., Lee, J., Fuchs, H.: Dynamic virtual convergence for video see-through head-mounted displays: Maintaining maximum stereo overlap throughout a close-range work space, New York City, International Symposium on Augmented Reality (ISAR) (2001)
4. Navab, N., Bani-Hashemi, A., Mitschke, M.: Merging visible and invisible: Two camera-augmented mobile c-arm (camc) applications. iwar **00** (1999) 134
5. Hofstein, S.: Ultrasound scope (April 1980) Patent 4200885, filed 2 Feb. 1978, granted 29 April 1980.
6. Masamune, K., Masutani, Y., Nakajima, S., Sakuma, I., Dohi, T., Iseki, H., Takakura, K.: Three-dimensional slice image overlay system with accurate depth perception for surgery. In: Medial Image Computing and Computer-Assisted Intervention (MICCAI). Volume 1935., Pittsburgh, Springer (2000) 395–402
7. Stetten, G.: System and method for location-merging of real-time tomographic slice images with human vision (July 2003) Patent 6599247, filed 11 Oct. 2000, granted 29 July 2003.

8. Drascic, D., Milgram, P.: Perceptual issues in augmented reality. In Fisher, S.S., Merrit, J.O., Bolas, M.T., eds.: Stereoscopic Displays and Virtual Reality Systems III. Volume 2653 of Proceedings of the SPIE., SPIE (1996) 123–124
9. Dodgson, N.A.: Autostereoscopic 3d displays. Computer **38**(8) (August 2005) 31–36
10. Tay, S., Blanche, P.A., Voorakaranam, R., Tunc, A.V., Lin, W., Rokutanda, S., Gu, T., Flores, D., Wang, P., Li, G., St Hilaire, P., Thomas, J., Norwood, R.A., Yamamoto, M., Peyghambarian, N.: An updatable holographic three-dimensional display. Nature **451**(7179) (2008) 694–698
11. Nowatzyk, A., Shelton, D., Galeotti, J., Stetten, G.: Extending the sonic flashlight to real time tomographic holography (2004) Workshop AMI-ARCS 2004, held in conjunction with MICCAI 2004 September 30th, 2004, Rennes (France). Published online at http://ami2004.loria.fr/.

Real-time Volume Rendering for High Quality Visualization in Augmented Reality

Oliver Kutter[1], André Aichert[1], Christoph Bichlmeier[1], Jörg Traub[1], Sandro Michael Heining[2], Ben Ockert[2], Ekkehard Euler[2], and Nassir Navab[1]

[1] Chair for Computer Aided Medical Procedures (CAMP)
Technische Universität München, Germany
{kutter,aichert,bichlmei,traub,navab}@cs.tum.edu

[2] Ludwig-Maximilians-Universität München, Chirugische Klinik und Poliklinik - Innenstadt AG CAS, Germany
{Sandro-Michael.Heining,Ben.Ockert,Ekkehard.Euler}@med.uni-muenchen.de

Abstract Recent advances in the field of computer graphics have enabled extremely fast and high quality rendering of volumetric data. However, these algorithms have been developed and optimized for visualization on single view displays, and not for stereoscopic augmented reality systems. In this paper, we present our implementation and results for the integration of a high quality hardware accelerated volume renderer into a medical augmented reality framework using a video see through HMD. The performance and quality of the renderer is evaluated through phantom experiments and an in-vivo experiment. Compared to the literature, our approach allows direct real-time stereo visualization of volumetric medical data on a HMD without prior time consuming pre-processing or segmentation. To further improve the visual perception and interaction of real and virtual objects, the renderer implements occlusion handling with the physicians hands and tracked medical instruments.

1 Introduction

Real-time visualization of 2D and 3D image data is one of the key tasks of any augmented reality system. Besides rendering the image data in real-time, the visualization has to provide the user with the appropriate visual cues at the right time. Augmented Reality was introduced as an alternative to monitor based visualization for the presentation of medical image data during surgery. Exemplary systems were proposed within neurosurgery [16], orthopaedic surgery [21], and liver surgery [9]. Existing solutions for medical augmented reality often implement real-time rendering by reducing quality or using simple rendering techniques. Simple wireframe surface models or slice rendering is applied to achieve real-time performance. In addition many methods require time consuming pre-processing of the volume data and manual interaction. More sophisticated methods enhance 3D perception within their stereo systems i.e. by transparent surfaces [6], virtual mirrors [3], or viewing windows [1] into the body. Important visual cues, such as shape from shading, depth from occlusion or motion parallax can help the user of a medical augmented reality system by an improved perception of the 3D data. This paper summarizes efficient implementations for hardware accelerated direct volume rendering (DVR) and proposes smart extensions to get the best visual perception while ensuring the required real time update rate.

2 Related Work

The development of GPU[3] ray-casting over the past five years [14,19] has led to flexible and fast hardware accelerated direct volume renderers. The popularity of GPU ray-casting is thanks to its easy and straightforward implementation on nowadays GPUs as well as the great flexibility and adaptability of the core ray-casting algorithm to a specific problem set.

Several recent works deal with enhancing the visual and spatial perception of the anatomy from the 3D data set by advanced DVR, many based on GPU ray-casting. Krüger *et al.*[13] propose ClearView for segmentation free, real-time, focus and context (F+C) visualization of 3D data. ClearView draws the attention of the viewer to the focus zone, in which the important features are embedded. The transparency of the surrounding context layers is modulated based on context layer curvatures properties or distance to the focus layer. Viola *et al.*[23] presented importance driven rendering of volumes based on cut aways allowing fully automatic focus and context renderings. Burns *et al.*[5] extend importance driven volume rendering with contextual cutaway views and combine 2D ultrasound and 3D CT image data based on importance. Beyer *et al.*[2] present a very sophisticated volume renderer, based on GPU ray-casting, for planning in neurosurgery. The renderer supports concurrent visualization of multimodal three-dimensional image, functional and segmented image data, skull peeling, and skin and bone incision removal at interactive frame rates. Kratz *et al.*[12] present their results for the integration of a hardware accelerated volume renderer into a virtual reality environment for medical diagnosis and education. The presented methods and techniques are closely related to our work. However the renderer is integrated within a virtual reality environment. Thus, no fusion of real and virtual data is performed and important augmented reality problems like interaction of real and virtual objects are not addressed.

F+C rendering for medical augmented reality, has been proposed in recent work by Bichlmeier *et al.*[4]. However was never fully realized using the GPU with the same degree of quality and performance as in the field of computer graphics. The presented methods are limited to rendering polygonal surface data, extracted by time consuming segmentation from the volume data set. Although the relevant anatomy and structures can be depicted, the approach suffers from lack of real volumetric visualization of medical image data. Important and standard medical visualization modes are impossible to achieve with these technique, i.e. Maximum Intensity Projection (MIP) and volume rendering of soft tissue. Another limitation is that the visualization is static and fixed to the time of mesh generation. Despite these drawbacks, the proposed video ghosting technique of the skin surface is promising as it enables a natural integration of the focus zone into the real video images.

Our work deals with the integration of a high quality hardware accelerated volume renderer into an augmented reality framework. The current setup is aimed at medical augmented reality using a video see through HMD system. Compared to surface rendering, DVR allows the use of the whole volume information during rendering without prior pre-processing.

[3] GPU - Graphics Processing Unit

3 Methods

3.1 Medical Augmented Reality System

The CAMPAR [20] augmented reality framework takes care of temporal synchronization of the video and tracking data streams and provides basic functionality for the visualization on the HMD. The video see-through HMD system is made of two color cameras for each eye and an inside-out optical tracking system which was originally introduced as RAMP by Sauer et al. [18]. CAMPAR runs on a standard workstation class PC (Dual Intel^{TM}Xeon, 3.2Ghz, 1.8Gb Ram, Nvidia^{TM}Geforce 8800 Ultra, 768Mb onboard memory). On the same PC runs also a server process for the RAMP inside-out tracking.

Tracking of the objects in the scene is accomplished by two separate optical tracking systems. An outside-in optical tracking system consisting of four infrared (IR) ARTtrack2 [4] cameras mounted to the ceiling of the room and an inside-out optical tracking system using an IR camera mounted directly on the HMD. Position and orientation of the users head are accurately determined by the inside-out tracking system, while the outside in tracking covers a much greater tracking volume and is used for all tracked objects in the scene. A reference system composed of a known arragement of optical tracking markers, visible by both tracking systems serves as the common reference frame of the scene.

For correct placement of the CT volume data inside the augmented reality scene it is registered to the coordinate frame of a tracking system in the scene. We use 4[mm] Beekley [5] CT-spots fiducials, that are additionally coated with an infrared reflective foil and attached to the surface of the subject. Thereby the same markers are locatable by the optical tracking system and in CT image data. The CT image and tracking space position of the markers are used to compute the rigid transformation from CT image to tracking space as described in [22]. Skin markers are however only suited for rigid anatomy, i.e. in phantom studies. For clinical introduction the concept of dynamic reference base i.e. markers attached to the target (bone) anatomy has to be used. If no combined markers or no markers at all were attached to the subject during CT image acquisition, the data has to be registered by other means. For phantom experiments (cf. Section 4) we therefore attached fiducials to the surface of a phantom and acquired a high resolution CT scan as the orginial human CT dataset was acquired without CT fiducials. The centroids of the markers are extracted fully automatically from the data set based on intensity thresholding and moments analysis. Finally the real human CT and phantom CT datasets are registered by intensity based registration, using *Mutual Information* and *Best Neighbor* optimization (cf. [11] for a comprehensive overview on medical image registration).

3.2 High Quality Volume Rendering

The GPU accelerated ray-casting based volume renderer is implemented as part of a C++ library in OpenGL and OpenGL's built-in shading language (GLSL). Rendering is realized by multipass techniques for ray setup, volumetric ray-casting, and composition of the final image. Various rendering modes, i.e. Non-Polygonal Iso Surface, Volume

[4] A.R.T. GmbH - http://www.ar-tracking.de/

[5] Beekley Corp. - http://www.beekley.com

Rendering, Maximum Intensity Projection (MIP), Iso Surface Shaded Volume Rendering and Digitally Reconstructed Radiographs (DRR), of 3D medical data are supported.

We employ several techiques to render high quality images in real-time. Floating point render targets are used for full precision throughout the entire rendering pipeline. Methods for clipping of the volume and for correct interaction of the volume and traditional opaque scene geometry similar to those presented by Scharsach *et al.*[19] are implemented. Local illumination is supported by various illumination models and on the fly gradient computation on the GPU. Fast and high quality non-polygonal iso surface rendering is realized by first hit iso surface ray-casting and subsequent iso surface refinement. Higher quality tranfer functions, pre-integrated [8] and post color attenuated [24], are employed to retain high image quality for low sampling rates. The performance of the renderer is optimized by empty space leaping of non visible voxels, using an octtree representation of the volume for culling of inactive voxels. Early ray termination in the ray-casting shaders, random ray start off-setting, and variable size of pipeline internal render targets futhermore contribute in achieving high performance and high quality rendering. For a detailed treatment of the rendering algorithms and techniques we refer to [7].

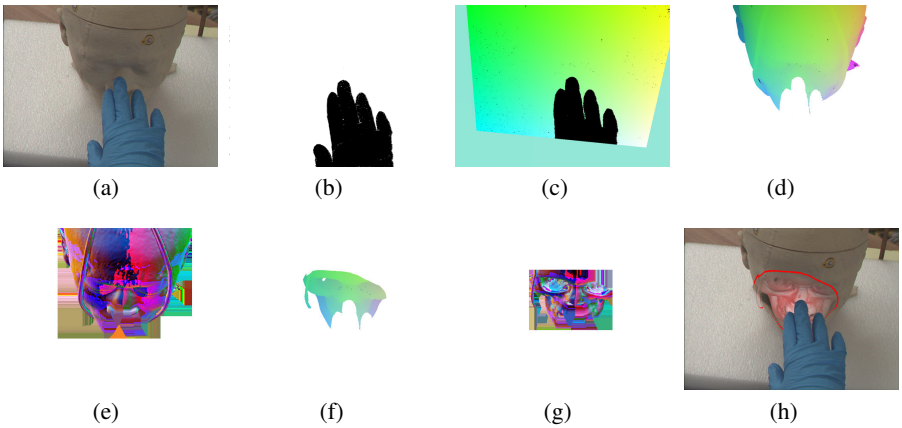


Figure 1. Stages of Focus and Context Rendering pipeline optimized for Augmented Reality with enabled Hand Occlusion Handling. (a) Video Input feed image from camera. (b) Depthbuffer for ray-casting after hand segmentation (c) Ray Entry Lookup texture (d,e) Skin Context Layer: (d) Surface, (e) Surface Normals. (f,g) Bone Focus Layer: (f) Surface,(g) Surface Normals. (h) Composition of Video Image, Focus and Context Layers.

In addition to the named standard volume rendering techniques the renderer implements ClearView [13] for pre-processing free, real-time F+C visualization (cf. Fig. 1(h)) of 3D data. The implementation is extended by the video ghosting techniques proposed by Bichlmeier *et al.*[4]. We furthermore add an optimization to the original ClearView pipeline greatly accelerating it for real-time stereo visualization. The original ClearView method generates focus and context layers each time the viewpoint changes without considering occlusion of inner layers by other closer to the camera located layers. Thereby, for a large number of pixels never contributing to the resulting image costly computations are executed. For an augmented reality environment, in which the viewpoint changes from frame to frame, this greatly reduces performance. We therefore incor-

porate focus region information already when generating the context and focus layers (cf. Fig 1(c) and 1(f)). Thereby, pixels outside the focus region are masked in an early stage of the pipeline and non-contributing computations are avoided. Especially more computationally expensive rendering techniques for inner layers benefit from this modifications. Additionally, the results of layer $n-1$ are used as inputs for layer n, when generating the temporary hit layer images.

3.3 Integration into AR environment

The requirements for volume rendering in medical augmented reality differ greatly from those for visualization on medical workstations, i.e. for diagnostic purposes. The rendering engine has to be optimized for speed first, followed by image quality second. The visualization should at no point result in large latencies of the overall system. In general this task is difficult to achieve as the performance depends on many parameters. Additionally, for in-situ visualization with a HMD as in our setup, the renderer has to be capable of perspectively correct stereo rendering in real-time (30fps). We realize stereo rendering, by rendering the volume for the left and right eye in two separate passes. In each pass the projection matrix for the current eye is loaded into the OpenGL projection matrix and the volume is rendered into a texture on the GPU. To achieve real-time stereo rendering performance we identified some basic optimization techniques that we applied when we adopted the volume renderer to the augmented reality environment.

A general guideline is to avoid computations that do not contribute to the final rendered image as early as possible in the pipeline. By making use of the occlusion of the volume by real and virtual objects further computations are skipped for many pixels of the image. The marking of pixels inside the focus zone for F+C rendering is motivated by the same considerations. Computationally intensive calculations should be deferred in the shaders as long as possible or skipped using conditional statements. A small memory footprint on the GPU is another important factor for high performance rendering. As much memory, textures, render and vertex buffers, should be shared or reused among different rendering techniques and for the per eye render passes. By keeping the memory usage to a minimum all data can be stored in high performance GPU memory. The image quality is another factor that has to be considered for the AR environment. Higher quality rendering techniques are computationally more costly resulting in a reduced framerate. The question is to which level of quality differences can be perceived by the user of the system on the HMD. Naturally, it makes no sense to render high resolution images if the same subjective user perception can be achieved with slightly reduced image resolution, resulting in a higher rendering performance.

Incorrect occlusion of virtual and real objects in the augmented scene is one of the fundamental problems in augmented reality. Without occlusion handling, virtual objects will appear floating over real objects and any impression of depth will be lost. In this work we present a basic yet effective method for occlusion handling in medical augmented reality using a video see through HMD. The operating site visible to the physician is relatively small and access to it is restricted to the treating physician, therefore the main and most disturbing occluders are the physician's own hands and instruments. In Sandor *et al.*[17] the advantages of hand occlusion handling for video see through HMDs were demonstrated. In our approach we handle the occlusion the physician hands by color filtering of the video images. A problem for the stable separation of a physician's hands from the rest of the video image is the large similarity of the standard gloves

white surface color and other features in the operating site image, i.e. patient skin. Therefore we exchange the standard white gloves by blue gloves, which are less common but available in the clinics. The color filter is implemented as a shader that compares color hue to a reference color where value and brightness are not too low for hue to be reliable. Because of natural shading, it is not sufficient to compare RGB color values. Brightness may vary largely, so conversion to HSV is worth the additional effort and gives much better results. The filter is applied at the very first stage of our rendering pipeline. Its result is written to a depth render target, where every pixel not belonging to one of the hands is set to the minimal depth value. The resulting depth buffer image (cf. Fig. 1(b)) is used for all following pipeline stages. Thereby, costly operations for pixels occluded by the hands are avoided by the Depth-Test of the GPU.

Fischer *et al.*[10], presented a method for instrument occlusion handling for medical augmented reality based on offline segmentation and first hit ray-casting. The integration and extension of their approach to real-time occlusion handling is straight forward within our method. It is implemented by running a depth test in a fragment shader, comparing the depth of the first hit isosurface, i.e. skin or bone, extracted by GPU ray-casting and the depth buffer values from the rendering of a polygonal model of a tracked instrument. For every pixel of the rendered instrument model with a smaller depth value than the skin iso-surface the pixel is set to black in the virtual instrument image. Thereby, after composition of all render pass results, the virtual instrument is only visible inside the patient behind the skin surface.

4 Experiments and Results

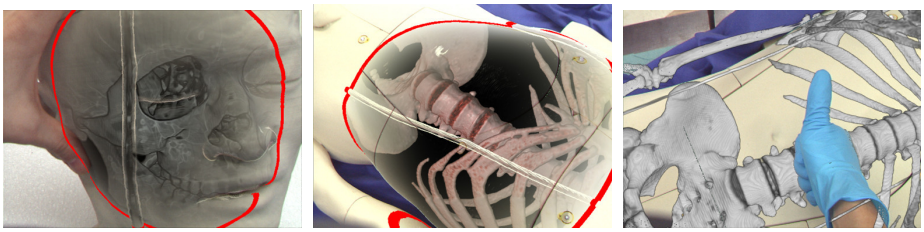
We conducted two experiments for human CT datasets on a realistic phantom and two in-vivo experiments to evaluate the performance of the renderer in a realistic scenario. The performance of the renderer was evaluated in detail for multiple rendering modes on the phantom. For the performance measurements the internal render target resolution was set to the camera resolution (640x480), the viewport resolution was set to the HMD resolution (1024x768). In all experiments empty space leaping and early ray termination optimizations were activated. For the evaluation a volunteer wore the HMD and inspected the phantom by continuous motion around it for two minutes. The minimum, maximum and, average framerate was measured using Fraps [6], a freely available benchmarking tool. A general note on the framerate before the detailed results are presented. The maximum framerate of the system is limited by the update rate of the slowest component in the system. For the used system the video cameras limit the overall system framerate. Thus, the average framerate can never be more than 30 frames per second, the camera refresh rate, in our experiments.

Phantom Experiments Setting up repeatable and realistic experiments for medical augmented reality is a tedious and complex task. A main factor is the availability of a realistic phantom. Within our group a phantom based on the data of the *Visible Korean Human Project* [15] [7] (VKHP) was manufactured [8]. The life-size phantom ranges from the head to the lower hips. The performance of the renderer was tested during several experiments with three datasets, head (296x320x420x8bit), thorax high res.

[6] http://www.fraps.com/

[7] http://vkh3.kisti.re.kr/new/

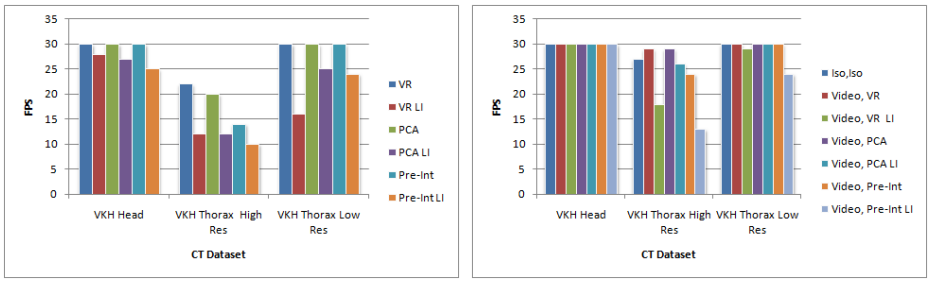
[8] FIT - Fruth Innovative Technologien GmbH, www.pro-fit.de



(a) (b) (c)

Figure 2. (a) Focus and Context rendering of the Korean head dataset. Video image is used as context layer, Focus layer is rendered with volume rendering. (b,c) Renderings of the Korean thorax dataset. (b) Focus and Context rendering with shaded volume rendering for the focus layer (bone). (c) Standard direct volume rendering visualizing the skeleton.

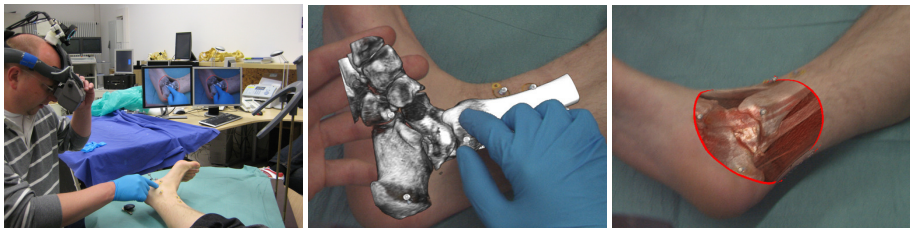
(256x368x522x8bit) and thorax low res. (256x368x261x8bit), extracted from the original *Visible Korean Human* (VKH) CT data. Examplary result images for the head and thorax datasets of the visible korean human phantom are shown in Fig. 2 (a-c). The measured average framerate for various DVR rendering modes is depcited in Fig 3(a) and Fig. 3(b) for F+C rendering. The better performance of the F+C rendering modes compared to the DVR rendering modes can be explained by the reduced number of pixels for which rays are cast through the dataset when the focus region optimization is enabled.



(a) Average FPS for various DVR modes (b) Average FPS for various F+C modes

Figure 3. Performance results for VKH datasets measured in average frames per second. (a) Average frames per second for various DVR modes. (b) Average frames per seconds for F+C Rendering. VR - Volume Rendering (VR) with standard post-classification, PCA - VR with Post Color Attenuation, Pre-Int - VR with Pre-Integration. LI - Local Illumination via Blinn-Phong and on-the-fly gradients by central differences. Iso - Skin Surface is rendered as shaded iso surface. Video - Ghosting of skin surface in video enabled.

In Vivo Experiment One of our colleagues, who had a sport accident, positioned CT markers at his right foot, using locations that could be identified later on for tracking markers, before a CT scan of the foot was acquired. By attaching tracking fiducials at these distinct landmarks the CT dataset with the CT fiducials at the same relative positions can be registered. In the experiment we evaluated the performance and quality of the rendered image for a volume (256x256x154 voxels), and as well the hand occlusion. Figure 4 depicts some of the results we got for the foot dataset with shaded DVR and enabled F+C rendering.



(a) (b) (c)

Figure 4. (a) Trauma surgeon inspecting the foot through the HMD. (b) Volume Rendering of bone, with visible Hand Occlusion problem. Hand without glove is beneath rendering. Hand with glove is placed correctly in front. (c) Focus and Context Rendering using shaded volume rendering for focus layer.

For the selected parameters the average framerate was 30fps. The frame rate never dropped below 28 fps, neither for standard VR nor for F+C rendering. We also evaluated the performance for rendering with the full HMD resolution (1024x768). Using this resolution, we noticed a slight performance drop of the average framerate for standard VR. For F+C rendering the framerate remained stable about 30fps. The performance drop for VR is mainly due to the full visibility of the dataset compared to F+C rendering, in which only the volume inside the focus zone is rendered. In the experiment the renderer was also tested by two members of the surgical staff of the trauma surgery department. After the experiments they provided us with valuable feedback considering performance, quality and possible future improvements. Especially the stable real-time performance and high quality of the renderer for several different anatomies, head, torso and the successful in-vivo experiment on the foot motivate further studies from the side of our clinical partners.

5 Discussion

Introduction of techniques for hardware accelerated volume rendering for medical augmented reality is an important step towards the integration of the complete 3D data into the medical augmented reality environment in real-time. The positive feedback of the physicians, motivates us for further investigation of the visualization and usability studies with multiple physicians.

The ultimate goal is to provide the best visualization of the 3D image data to the physicians at any point during a medical procedure. However, functionality comes before fancy visualization.Thus, determining the best rendering mode for each medical application is another important issue and interesting subject of research. Future work should focus on evaluation of the already existing rendering modes for a defined set of medical procedures together with our clinical partners. The specific requirements have to be analyzed for each procedure over the complete workflow, from pre-operative imaging till the end of the intervention. For this purpose we plan to record video and tracking data for exemplary procedures for use as offline training datasets. For these dataset the visualization modes can be interactively changed and adjusted based on feedback from our medical partners. Thereby, the visualization can be iteratively refined and optimized for the specific medical procedure.

Concerning the currently implemented occlusion handling, we do not claim that the method is suited for augmentation of all medical procedures, especially not for traditional open surgery. Simple color segmentation is not enough for this task. However, for pre-operative inspection of the patient using augmented reality techniques through the HMD or as mentioned for minimally invasive procedures, with access restricted by ports into the patient, the approach is sufficient. The integration of more elaborate methods for real-time occlusion handling, i.e. depth map reconstruction using the stereo camera system, could be seen as future work.

6 Conclusion

This papers presents and summarizes the results for the integration of a high quality hardware accelerated volume renderer into an augmented reality environment. The pipeline of the renderer and advanced rendering techniques were adapted and optimized for stereoscopic rendering on a HMD. The occlusion problem of real and virtual objects in augmented reality scenes was addressed in this work, by simple yet effective methods for diagnosis and minimally invasive procedures using a HMD. The method allows an improved, more natural perception of the augmented reality scene. The performance of the presented techniques has been demonstrated and evaluated in several experiments by rendering real human CT data of the skull and the complete thorax on realistic anatomical phantoms using different rendering techniques. In an in-vivo experiment the renderer was used to augment a human foot by its pre-acquired CT dataset and evaluated by three trauma surgeons.

References

1. M. Bajura, H. Fuchs, and R. Ohbuchi. Merging virtual objects with the real world: seeing ultrasound imagery within the patient. In *Proceedings of the 19th annual conference on Computer graphics and interactive techniques*, pages 203–210. ACM Press, 1992.
2. J. Beyer, M. Hadwiger, S. Wolfsberger, and K. Bühler. High-Quality Multimodal Volume Rendering for Preoperative Planning of Neurosurgical Interventions. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, pages 1696–1703, 2007.
3. C. Bichlmeier, M. Rustaee, and S. H. nad N. Navab. Virtually extended surgical drilling device: Virtual mirror for navigated spine surgery. In *Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2007.
4. C. Bichlmeier, F. Wimmer, H. Sandro Michael, and N. Nassir. Contextual Anatomic Mimesis: Hybrid In-Situ Visualization Method for Improving Multi-Sensory Depth Perception in Medical Augmented Reality. In *Proceedings of the 6th International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 129–138, Nov. 2007.
5. M. Burns, M. Haidacher, W. Wein, I. Viola, and E. Groeller. Feature emphasis and contextual cutaways for multimodal medical visualization. In *EuroVis 2007 Proceedings*, May 2007.
6. P. J. Edwards, L. G. Johnson, D. J. Hawkes, M. R. Fenlon, A. J. Strong, and M. J. Gleeson. Clinical experience and perception in stereo augmented reality surgical navigation. In *Proceesings of Medical Imaging and Augmented Reality: Second International Workshop, MIAR 2004, Beijing, China, August 19-20, 2004.*, pages 369–376, 2004.
7. K. Engel, M. Hadwiger, J. Kniss, and C. Rezk-Salama. *Real-Time Volume Graphics*. AK Peters, Ltd., 2006.
8. K. Engel, M. Kraus, and T. Ertl. High-quality pre-integrated volume rendering using hardware-accelerated pixel shading. *Proceedings of the ACM SIGGRAPH/EUROGRAPHICS workshop on Graphics hardware*, pages 9–16, 2001.

9. M. Feuerstein, T. Mussack, S. M. Heining, and N. Navab. Intraoperative laparoscope augmentation for port placement and resection planning in minimally invasive liver resection. *IEEE Trans. Med. Imag.*, 27(3):355–369, March 2008.
10. J. Fischer, D. Bartz, and W. Straßer. Occlusion handling for medical augmented reality using a volumetric phantom model. *Proceedings of the ACM symposium on Virtual reality software and technology*, pages 174–177, 2004.
11. J. Hajnal, D. Hawkes, and D. Hill. *Medical Image Registration*. CRC Press, 2001.
12. A. Kratz, M. Hadwiger, A. Fuhrmann, R. Splechtna, and K. Buhler. GPU-Based High-Quality Volume Rendering For Virtual Environments. In *AMIARCS*, Copenhagen, Denmark, Oct. 2006. MICCAI Society.
13. J. Krüger, J. Schneider, and R. Westermann. ClearView: An interactive context preserving hotspot visualization technique. *IEEE Transactions on Visualization and Computer Graphics (Proceedings Visualization / Information Visualization 2006)*, 12(5), September-October 2006.
14. J. Krüger and R. Westermann. Acceleration Techniques for GPU-based Volume Rendering. *Proceedings of the 14th IEEE Visualization 2003 (VIS'03)*, 2003.
15. J. Park, M. Chung, S. Hwang, Y. Lee, D. Har, and H. Park. Visible korean human: Improved serially sectioned images of the entire body. 24(3):352–360, March 2005.
16. P. Paul, O. Fleig, and P. Jannin. Augmented virtuality based on stereoscopic reconstruction in multimodal image-guided neurosurgery: Methods and performance evaluation. *IEEE Trans. Med. Imag.*, 24(11):1500–1511, 2005.
17. C. Sandor, S. Uchiyama, and H. Yamamoto. Visuo-Haptic Systems: Half-Mirrors Considered Harmful. *EuroHaptics Conference, 2007 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics 2007. Second Joint*, pages 292–297, 2007.
18. F. Sauer, A. Khamene, B. Bascle, and G. J. Rubino. A head-mounted display system for augmented reality image guidance: Towards clinical evaluation for imri-guided neurosurgery. In *Proc. Int'l Conf. Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pages 707–716, London, UK, 2001. Springer-Verlag.
19. H. Scharsach, M. Hadwiger, A. Neubauer, S. Wolfsberger, and K. Buhler. Perspective Isosurface and Direct Volume Rendering for Virtual Endoscopy Applications. *Proceedings of Eurovis/IEEE-VGTC Symposium on Visualization 2006*, pages 315–322, 2006.
20. T. Sielhorst, M. Feuerstein, J. Traub, O. Kutter, and N. Navab. Campar: A software framework guaranteeing quality for medical augmented reality. *International Journal of Computer Assisted Radiology and Surgery*, 1(Supplement 1):29–30, June 2006.
21. J. Traub, P. Stefan, S.-M. M. Heining, T. Sielhorst, C. Riquarts, E. Euler, and N. Navab. Towards a hybrid navigation interface: Comparison of a slice based navigation system with in-situ visualization. In *Proceedings of MIAR 2006*, LNCS, pages 364–372, Shanghai, China, Aug. 2006. Springer.
22. J. Traub, P. Stefan, S.-M. M. Heining, C. R. Tobias Sielhorst, E. Euler, and N. Navab. Hybrid navigation interface for orthopedic and trauma surgery. In *Proceedings of MICCAI 2006*, LNCS, pages 373–380, Copenhagen, Denmark, Oct. 2006. MICCAI Society, Springer.
23. I. Viola, A. Kanitsar, and M. Groller. Importance-driven volume rendering. *Proceedings of IEEE Visualization*, 4:139–145, 2004.
24. Q. Zhang, R. Eagleson, and T. Peters. Rapid Voxel Classification Methodology for Interactive 3D Medical Image Visualization. In *MICCAI 2007 Proceedings*, Lecture Notes in Computer Science. Springer, Oct. 2007.

Tracking–based statistical correction for radio–guided cancer surgery

Alexander Hartl[1], Thomas Wendler[1], Tobias Lasser[1], Sibylle I. Ziegler[2], and Nassir Navab[1]

[1] Computer Aided Medical Procedures (CAMP), TUM, Munich, Germany,
[2] Nuclear Medicine Department, Klinikum rechts der Isar, TUM, Munich Germany

Abstract. Radio–guided control of tumor resection borders can be provided by intra-operative hand–held beta–probes. As radioactive decay and its detection follow a statistical process the readings detected with such probes suffer from a high variance over time. In general their measurements are averaged over a certain duration, e.g. $1-3$ $[s]$, in order to provide statistically reliable measurements. This results in a smearing of the readings over an area if the probe is moved and thus in the risk that microscopic cancer residuals remain undetected.
To deal with this problem a statistical correction based on tracking of the hand–held detector is introduced. By extending the probe with a tracking system, statistical correction of the readings with a position–filtering approach is possible. In addition the new system allows calculating the confidence level of the estimated probe measurements in the region of interest, which is displayed interactively. This enables guiding the surgeon during the acquisition, making the process less user dependent and more robust.
The results do not only prove the concept, but show the superiority of the novel approach. A correlation with the ground truth of $53-75\%$ versus $38-51\%$ of the standard procedure was achieved. We show in this paper how computer assistance contributes to advanced visualization and guidance in surgery, as well as how it supports improving the quality of the acquired data.

1 Introduction

Lately intra–operative nuclear medicine guided procedures have been introduced in cancer surgery in order to allow radio–guided control of tumor resection borders. In this procedure, a radioactive tracer that accumulates primarily in tumors cells (i.e. FDG) is injected into the patient before the intervention. After the main tumor is resected with minimal resection margins, a hand–held radiation detector, i.e. a beta–probe, is used to find microscopic cancer residuals in the tumor bed, as these present higher count rates over the background radiation [1, 2]. This procedure allows the surgeon to selectively expand the resection borders.

The acquisition of radiation readings is a statistical process due to the random nature of the radioactive decay. In standard procedures this results in noisy or

even invalid data. Moreover it forces the use of temporal smoothing, like moving average filters with durations of $1-3$ $[s]$. Unfortunately, the filtering introduces another pitfall for moving, hand–held probes: small radioactive 'hot spots' are smeared out and thus can vanish amidst the background radiation.

The combination of beta–probes with tracking systems was introduced for computer–assistance in cancer resection procedures for advanced visualization and instrument guidance [3]. By scanning the resection borders with synchronous recording of position, orientation and radio–active beta–probe measurements, they reconstructed surfaces showing an activity distribution.

To the authors' knowledge, there exists no work that compensates for the statistical variance of the readings, which results in noisy and smeared surface measurements. In this paper we introduce new concepts for a computer–assisted intervention system by statistically correcting the readings of a tracked beta–probe using maximum likelyhood estimators and spatial information rather than filtering over time.

The statistical compensation is used to implement a computer–assisted guidance of the acquisition process, which results in a less user dependent and more robust procedure.

2 Materials and Methods

2.1 System setup

A hand–held beta–probe (Silicon Instruments GmbH, Berlin, Germany) was combined with an infrared optical tracking system in order to acquire the position and orientation of the probe synchronously with its readings (cf. figure 1). The tracking system consists of four ARTtrack2 infrared tracking cameras (Advanced Realtime Tracking GmbH, Weilheim, Germany). The cameras are located so that they are able to track a volume of $2 \times 2 \times 2$ $[m^3]$ in which they achieve a root mean square (rms) error of 0.4 $[mm]$ for the target position and 0.002 $[rad]$ for the orientation according to the manufacturer's specification (maxima 1.4 $[mm]$ and 0.007 $[rad]$, respectively).

2.2 Probe calibration

The calibration of the beta–probe consists of two steps. In a first step, the rigid transformation from the attached tracking target to the detector is determined. A linear 5DOF tracking target is used that is aligned with the axis of the probe. This is the distance between the tracking target and the detector along the axis of the probe.

In a second step, the field of view of the probe is obtained by acquiring the measurements of the beta–probe due to a single point source of radiation located at different, well defined positions. For the source, a cellulose scaffold of 2×2 $[mm^2]$ with one 'hot spot' of F^{18}–2–fluoro–2–deoxy–D–glucose (FDG, 1 $[MBq/ml]$) is placed on a 3–axes positioning table. The measurements of the

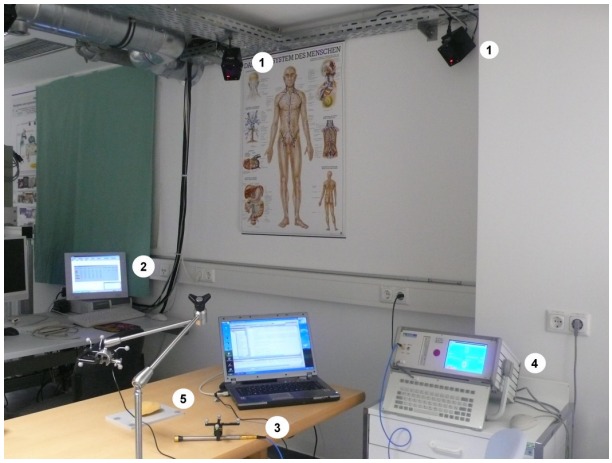


Fig. 1. The system setup consists of (1) tracking cameras, (2) tracking computer, (3) tracked beta–probe, (4) probe computer, and (5) radioactive phantom.

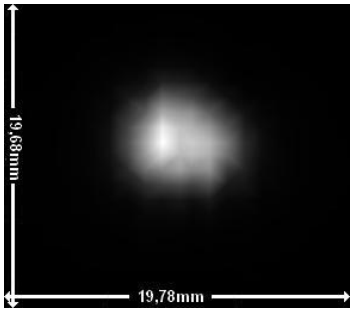
beta–probe at each distinct position with step size of 0.1 $[mm]$ are averaged over 10 $[s]$ to ensure a certain statistical reliability (fixed height of 1 $[mm]$ over source). The resulting uniformly spaced data–set (cf. figure 2(a)) is then analyzed to determine the parameters of the field of view of the beta–probe (cf. figure 2(b)).

2.3 Statistical correction and confidence calculation

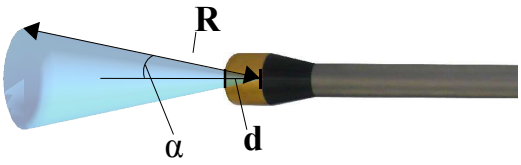
For the position–based statistical correction and confidence calculation a 2D grid is interactively defined by the user on the region of interest. A surface reconstruction with a tracked infrared laser pointer is performed before that (as proposed in [4], mean quadratic error is 1.8 ± 0.4 $[mm]$) to enable projection of the grid onto the phantom with correct height information.

The relative position and orientation of the tip of the probe to the grid is used to calculate the contribution of each probe measurement to each distinct grid pixel (cf. figure 3(a)). This is achieved by defining a virtual camera with the same properties as the probe and then rendering the scene from the view of the probe. We use the center of projection, the field of view, and the maximum range of detection as the 'far' plane of a standard OpenGL camera within the rendering pipeline. The reading acquired and the contribution of each grid cell to the reading is then stored for further processing. The contribution to a certain grid position is defined as the probability that the current reading was originated solely because of the said grid position. An ad–hoc model of the acquisition process is required, see [5] for more details.

In order to correct for the statistics all contributions and probe readings for each grid position are given an optimal estimator. In case of a Poisson distribution, which is a standard model for both the decay and detection process of



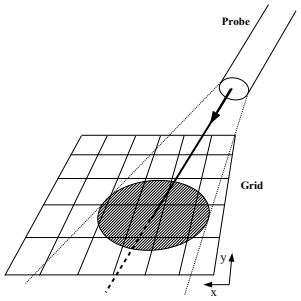
(a) Intensity map of probe calibration measurements of a single F^{18} point source at different positions (step size 0.1mm).



(b) Parameters to model the field of view of the beta–probe: angle of aperture α, center of projection d, maximum range of detection R.

Fig. 2. Experiment and model for beta–probe calibration.

radioactivity with high decay rates, the maximum likelihood estimator is the average of all readings. The confidence of the estimation at each grid point is also calculated using the readings obtained so far in an algorithm based on Pearson's χ^2-goodness-of-fit test [5]. This value represents the probability that the measurements are related to a Poisson process with their mean measurement as expected value.



(a) The readings of the probe are assigned to the grid cells that are within the field of view.

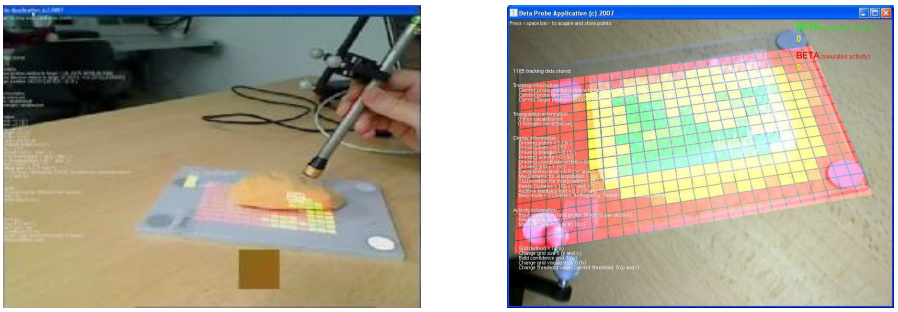
(b) Exemplary grid structure projected onto a phantom surface.

Fig. 3. A discrete grid is defined and mapped to the surface. The measurements of the probe at each distinct position contribute to several pixels within the grid.

2.4 Computer–assisted acquisition

The availability of the confidence enables guidance of the scanning process. This solves the current problem of not being able to guarantee that the control of the resection borders was thorough enough and that no spot was not controlled for sufficient time to have a statistically valid reading.

In this work the guidance of the surgeon during the acquisition of the data was achieved by calculating the confidence at all grid points in real time and visualizing its value in every grid cell by means of augmented reality (figure 4(b)). The value was encoded with a color in the range from red to green, where green indicates a high level of confidence and red a low level of confidence. These results in an intuitive way to exhibit which areas require additional measurements.



(a) Guided beta–probe scan process. (b) Incomplete beta–probe scan.

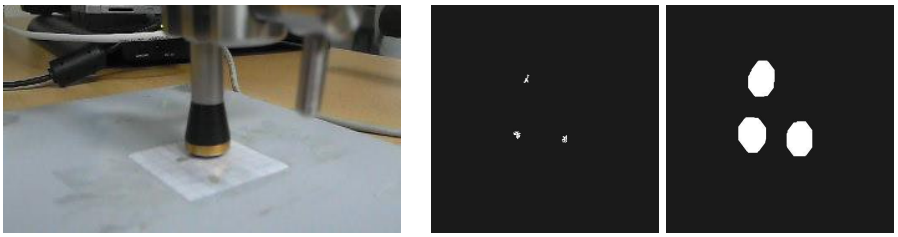
Fig. 4. Visualization of confidence values during the scan (a) and at the end of a scan procedure (b). Confidence values are overlaid onto an image of a video camera in real time. Green indicates cells with a high confidence level, red indicates a low confidence level and recommends an additional scan within these areas.

3 Experiments and results

For the experiments two flat cellulose phantoms with three 'hot spots' each (100 $[kBq/ml]$) were used, one with a background radioactivity (10 $[kBq/ml]$) and the other without background radiation. The confidence grid was fixed to the corners of the phantoms.

The 'hot spots' were colorized differently, enabling the ground truth to be obtained by scanning the phantom with an office scanner and segmenting them using a region-growing approach. The ground truth image was then registered to the grid using a 2D point-based registration (reprojection mean error 0.03 $[px]$).

Furthermore, the image of the ground truth was dilated by a morphological operator defined from the field of view of the probe. This was meant to simulate the fact that the probe is not a point measurement and that the sensitivity is almost constant in a radius of 3.11 $[mm]$ (figure 2(a)).



(a) Experiment. (b) Ground truth.

Fig. 5. (a) A cellulose phantom with three 'hot spots'. (b) Segmented and dilated 'hot spots'.

For both phantoms seven sets of measurements were obtained by different test persons. The visualization of the confidence was used in all cases. The number of measurements was 1513 ± 107 for a duration of 215 ± 35 $[s]$. The pixel size of the grid was set to be 2.5 $[mm]$ for a 3×3 $[cm^2]$ area.

For each set reconstructions of the surface radioactivity were performed using directly the raw data, a 1 $[s]$ moving average time filter (equivalent to clinical standard procedure) and the position–based statistical correction. All three reconstructions were then compared with the ground truth.

In all reconstructions with and without background radiation the 'hot spots' could be clearly identified when statistical filtering was used (cf. figure 6). Furthermore the images generated were very smooth and did not present the noise patterns that were reported in the literature [3]. With the standard time filtering approach an identification of the 'hot spots' in the reconstructed image was difficult, in particular in presence of background activity.

For a quantitative comparison the normalized cross correlation (NCC) and normalized mutual information (NMI) between the reconstructed surfaces and the ground truth was computed. In all reconstructions the position–based statistical filtering was superior with both NCC and NMI.

4 Discussion

In this work the idea of using tracking and statistical models for the correction of beta–probe readings was introduced. The concept of computer–guided scanning was presented for the first time. The results show that such a computer–assistance can reduce the risk of missing microscopic cancer residuals in procedures like beta–probe control of resection borders. The use of this information for the reconstruction of radioactivity surfaces was proved to be statistically more robust.

There are, however, several considerations for this new approach. The first point is the implicit assumption that the region to be scanned does neither move nor deform during the scan. Movement and deformation of anatomy after the

Without background

	NCC			NMI		
	statistical filter	raw data	time filter	statistical filter	raw data	time filter
mean	0.76	0.51	0.35	0.86	0.98	0.99
std	0.05	0.07	0.08	0.03	0.0	0.01
min	0.70	0.40	0.24	0.83	0.97	0.98
max	0.82	0.58	0.46	0.9	0.99	0.99

With background

	NCC			NMI		
	statistical filter	raw data	time filter	statistical filter	raw data	time filter
mean	0.53	0.38	0.28	0.83	0.97	0.99
std	0.08	0.08	0.04	0.09	0.01	0.00
min	0.44	0.29	0.23	0.68	0.96	0.98
max	0.65	0.47	0.35	0.97	0.99	0.99

Table 1. The normalized cross correlation (NCC) and normalized mutual information (NMI) between the reconstructed images and the ground truth for the phantom without background radiation.

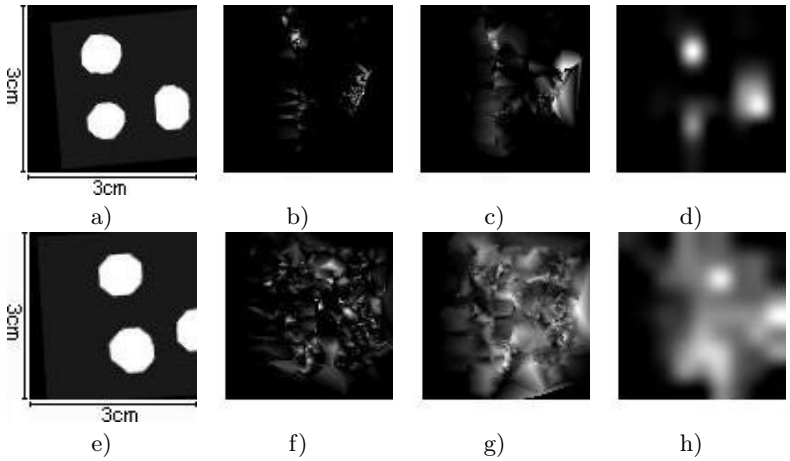


Fig. 6. The images computed from the measurements on the phantom without background radiation (a-d) and with background radiation (e-h). From left to right: ground truth, raw data, time filtering, and statistical filtering.

scan is unproblematic, as the generated images are snapshots of the current situation. Due to the short acquisition time acquiring a new scan is easily possible and the information can be updated when needed. Our group is currently working on these issues, for example detection of deformation and movement that could invalidate the generated images.

Concerns of bringing further equipment to the operation room are nowadays less relevant. In the case of tracking for example, tailored solutions for surgery are commercially available. Simultaneously, costs of such surgical solutions have been continuously decreasing. Furthermore there might be radiation safety issues for the surgical personnel, but as shown by Piert et al. for FDG radio–guided surgery the doses are negligible [6].

Regarding the tracer to be used, it was taken for granted that the tumor residual has an uptake that allows its detection over the background. This is indeed not always the case. In our institution we have thus developed criteria for selecting proper patients, based on the analysis of pre–operative PET imaging. Related to this, a current issue of research is the minimal amount of residual cancer cells that can be detected with a beta–probe. Daghighian et al report that the beta–probe used by them can detect tumor residuals down to 50 $[mg]$ [7]. Updating these limits are subject of research in our group.

As future work we see special potential in including global reconstruction of the radioactivity distribution rather than only doing local statistical corrections targeting higher quality surface images.

In summary, this paper introduces novel ideas for significantly improving the robustness and the statistical quality of radioactive measurements, and thus improving radio–guided resection border control. The proof–of–concept shows that computer–assistance does not only incorporate advanced visualization and guidance into intra–operative procedures, but also supports the process of reliable and robust functional image acquisition.

References

1. Franc, B.L., Mari, C., Johnson, D., Leong, S.P.: The role of a positron- and high-energy gamma photon probe in intraoperative localization of recurrent melanoma. Clin Nucl Med **30**(12) (2005) 787–791
2. Zervos, E.E., Desai, D.C., DePalatis, L.R., Soble, D., Martin, E.W.: 18f-labeled fluorodeoxyglucose positron emission tomography-guided surgery for recurrent colorectal cancer: a feasibility study. J Surg Res **97**(1) (2001) 9–13
3. Wendler, T., Traub, J., Ziegler, S.I., Navab, N.: Navigated three dimensional beta probe for optimal cancer resection. In: Proc. MICCAI 2006. Volume 1 of LNCS 4190., Copenhagen, Denmark, MICCAI Society, Springer (2006) 561–569
4. : -. In: -. (2008) –
5. : -. PhD thesis (2008)
6. Piert, M., Burian, M., Meisetschlger, G., Stein, H.J., Ziegler, S., Nhrig, J., Picchio, M., Buck, A., Siewert, J.R., Schwaiger, M.: Positron detection for the intraoperative localisation of cancer deposits. Eur J Nucl Med Mol Imaging **34**(10) (2007) 1534–1544
7. Daghighian, F., Mazziotta, J.C., Hoffman, E.J., Shenderov, P., Eshaghian, B., Siegel, S., Phelps, M.E.: Intraoperative beta probe: a device for detecting tissue labeled with positron or electron emitting isotopes during surgery. Med Phys **21**(1) (1994) 153–157

Registration of a 4D Cardiac Motion Model to Endoscopic Video for Augmented Reality Image Guidance of Robotic Coronary Artery Bypass

Michael Figl[1,6], Daniel Rueckert[1], David Hawkes[3], Roberto Casula[4], Mingxing Hu[3], Ose Pedro[1], Dong Ping Zhang[1], Graeme Penney[5], Fernando Bello[2], and Philip Edwards[1,2]

[1] Department of Computing, Imperial College London, UK
[2] Department of Biosurgery and Surgical Technology, Imperial College London, UK
[3] Centre of Medical Image Computing, University College London, UK
[4] Cardiothoracic Surgery, St. Mary's Hospital, London, UK
[5] Division of Imaging Sciences, King's College London, UK
[6] Center for Biomedical Engineering, Medical University of Vienna, Austria

Abstract. The aim of the work described in this paper is registration of a 4D preoperative motion model of the heart to the video view of the patient through the intraoperative endoscope, in order to overlay the real video sequence with it. As the heart motion is cyclical it can be modelled using multiple reconstructions of cardiac gated coronary CT.
We propose the use of photoconsistency between the two views through the da Vinci endoscope to align to the preoperative heart surface model from CT. We propose averaging of the photoconsistency over the cardiac cycle to improve the registration compared to a single view.
Results are presented for simulated renderings and for real video of a beating heart phantom. We found much smoother behaviour of the test function at the minimum when using multiple phases for the registration, furthermore convergence was found to be better when more phases are used.

1 Introduction

Totally endoscopic coronary artery bypass (TECAB) is a minimally invasive robotic cardiac procedure performed on the beating heart. The view of the patient is provided by a stereo endoscope which gives the surgeon a 3D view of the operative field.

There is a significant proportion of these operations that need to convert to more conventional procedures. This may be due to misidentification of the diseased vessel or difficulty in locating the vessel through fatty deposits on the surface of the myocardium. For this reason there has been some interest in introducing augmented reality (AR) image guidance to TECAB, overlaying these vessels with some preoperative images. Mourgues et al. [6] and Adhami et al [1] proposed such a system where the preoperative model to be overlaid over the real image comes from bilateral x-rays.

In our case the preoperative model comes from coronary CT. This provides a fully 4D model of the heart from which a motion model of both the myocardium and coronary arteries must be extracted.

To overlay the endoscopic view the preoperative model has to be registered to the endoscopic images. If many landmarks are visible, e.g. bifurcations or tubes, we could use a shape-based registration such as that proposed by Jomier [5]. However in our case only parts of the heart surface are seen via the endoscopic camera and there may not be corresponding vessels visible on the heart surface and in the preoperative image.

Since we have a stereo video stream it is possible to use photo-consistency [3] as a measure of registration to the preoperative surface. If we have temporal synchronisation between the video and the motion model, the periodic motion of the heart enables us to use all points of time in the cardiac cycle.

2 Materials & Methods

2.1 Heart Phantom

To allow for reproducible testing the motion model was derived from a 4D CT scan of a pneumatic heart phantom made by The Chamberlain Group, Great Barrington, MA, USA. This phantom allows three different heart frequency settings. The heart rate of the phantom was determined using a video sequence of the beating phantom and comparing the first image of this sequence to the others using cross-correlation as a similarity measure. The frequency was found as a peak in the Fourier transform of this function.

The phantom was scanned by a Philips Brilliance 64 ECG triggered CT, and reconstructed at ten different phases of the cardiac cycle. The ECG signal for the heart phantom was generated by a signal generator from Hameg, Mainhausen, Germany. The surfaces of the ten phases were extracted using the VTK implementation of Marching Cubes (Kitware Inc, NY, USA).

2.2 Video Images

The target of our work is the daVinci robot which is equipped with a stereo endoscope with stereo video output via BNC and S-video. As a development system two Sony DCR-TVR 30E digital video cameras were used. Video grabbing was performed using an LFG4 4 channel frame grabber made by Active Silicon, Uxbridge, UK. The video cameras were calibrated using a calibration toolbox implemented in `MatLAB` by Jean-Yves Bouguet, [2].

2.3 Temporal Registration

Aligning the phases of the video sequence and the 4D CT is done by visual judgment. There are several time points that can easily be found in both the rendered CT surface and the video sequences and can be used for temporal synchronisation. E.g. the end of the systolic phase, where the heart is in maximal contraction. In practice this could also be achieved using the ECG signal

2.4 Point Based Registration

In a first step a simple and robust point based registration of the surface to the video images is done. Several points on the surface and in both video channels are selected, and the optimal rigid body transformation T is found by minimising the distance of the projected surface points to the selected image points iteratively. This registration software was developed using the OpenGL widget of the C++ GUI framework Qt3/4 (Trolltech, Oslo, Norway). Multidimensional minimisation was done using the Fletcher-Reeves conjugate gradient algorithm implemented in the GNU scientific library.

2.5 Photo-consistency Registration

Point based registration was used as a starting value for photoconsistency, see figure 1.

Fig. 1. Registration using photoconsistency. In the upper image the situation after the point to point registration step is shown. The bottom image shows the effect of application of photoconsistency registration.

The idea behind registration using photo-consistency is to judge whether given projections of a surface are compatible with the lighting model. Assuming a Lambertian lighting model [4] a surface point should have the same colour in all of its projections. Figure 2 gives an example, where three points have similar colour values in different images. The mean standard deviation is 2.3 in an 8 bit colour space, but the standard deviation for the green component in the three points of the second image is 26.2.

For the registration we have to find the transformation from the surface to the common coordinate system of the camera calibrations. After applying this

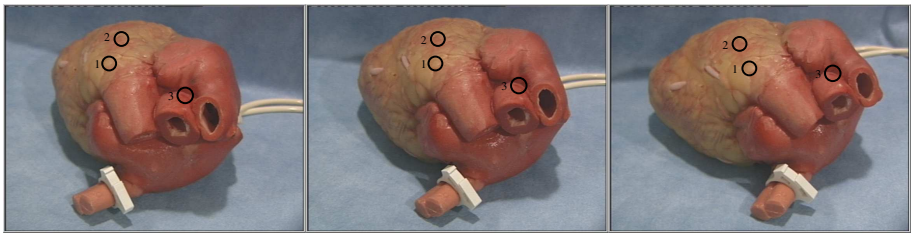


Fig. 2. With Lambert's light model the colours in the centers of the circles 1, 2, and 3 should be the same in every image. The mean standard deviation of colours values in these points is 2.3 in an 8 bit colour space. On the lower front of the second and third image there is a considerable reflection which does not comply with the Lambertian model and has a very different colour in the first image.

transformation the surface object will be in a position where the projections of the surface points are (photo-)consistent. This consistency can be evaluated using the colour components of the projections of the surface points in different video channels. If the colour diversity in the images is evaluated by their variance, then the desired transformation T can be found by:

$$T = \operatorname{argmin} \left(\operatorname*{mean}_{x} (\operatorname*{var}_{i} (P_i T x)) \right) \tag{1}$$

Where P_i denotes the projection matrix for camera calibration i. The comparison for the mean value is done over the surface points that are visible in both images. Figure 3 shows the situation for three cameras. Photo-consistency as a similarity measure for 2D-3D registration was introduced by Clarkson et al. [3].

2.6 Photo-consistency with an image sequence

To take advantage of the fact that there is a whole image sequence available for registration, one can try to use the different phases together to get the registration. Assuming that all the deformation or movement of the object can be found in the surfaces and is completely periodic, the transformation from the preoperative to the intraoperative coordinate system is always the same. In figure 4 the chain of transformations from the preoperative coordinate system to the display is shown for three different phases. Both the transformation T and the camera calibrations are unchanged in this idealised situation. This assumption was also used for 3D-3D registration using the iterative closest point algorithm as proposed by Wilson et al. [8].

3 Results

For camera calibration we used 12 images of a checkerboard pattern with 11×9 squares for every channel. The baselines of the squares was 7.5mm. Typical calibration errors with the endoscope were 0.51 and 0.45 pixels in x resp. y direction

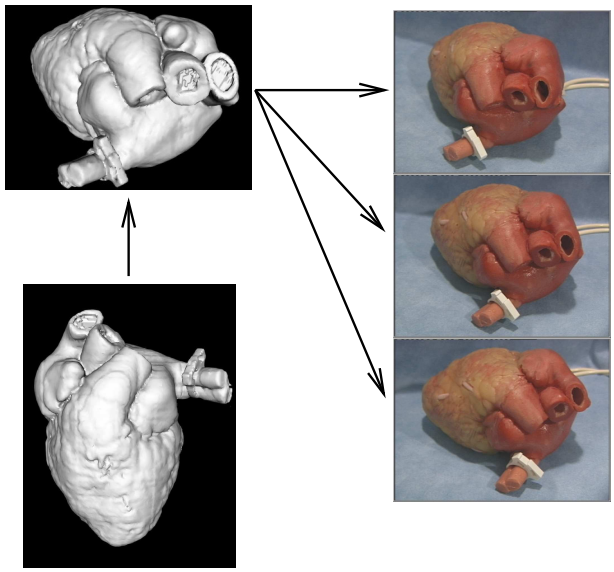


Fig. 3. The transformation T has to move the surface object in a way that the projections of every point of the surface into the video images using the camera calibrations P_i are photo-consistent.

for the separate channels before undistortion and stereo calibration. After undistortion and stereo calibration the distortion coefficients were ten times smaller and the errors in the same magnitude as before, in a concrete example 0.4, 0.33 pixels.

The heart rate could readily be determined using the method described in 2.1, both for endoscopic videos of the beating heart phantom and of a real patient's heart. The frequency measurements on the phantom could be verified by simple counting, the heart rate measured on the patient was identical to the ECG measured heart rate.

3.1 Renderings as Target Images

To evaluate the method without introducing errors by the camera calibration, the temporal alignment or the quality of the video images we did a first registration series from the surface model to renderings of the surfaces. The renderings were made using camera calibration data from the robot's stereo endoscope neglecting the distortion. Unsurprisingly the graph of the photo-consistency surface is smoother when more phases are used. To quantify this, the total variation (normalised by the range) of the graph along the coordinate axes is derived:

$$T(f) := \frac{1}{\max(f)-\min(f)} \int \|f'\| \tag{2}$$

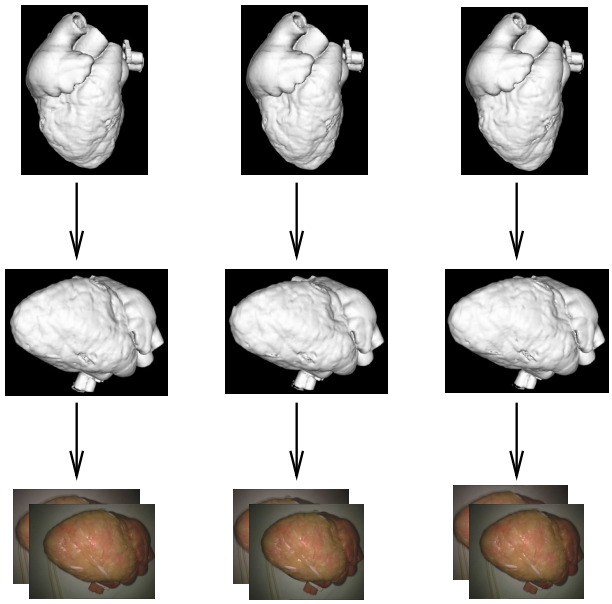


Fig. 4. The transformation between the preoperativly generated surfaces and the intraoperative position is the same for all phases, provided there is no camera movement or respiration.

The corresponding normalised total variations for the function in figure 6 were 1.69, 1.43, and 1.4. The remaining variations can be found in table 5(a).

3.2 Video Cameras

Using the development system consisting of two Sony video cameras we did measurements comparable to that were done with the renderings.

In this test series up to ten phases of the heart cycle were used. The results for translation along the z-axis are shown in figure 6. In figure 6 the situation at the minimum for translation along the x-axis of the world coordinate system is shown when using one, five, or ten phases.

Furthermore we made a test series to compare the convergence, starting at a 5 different positions. The starting positions were found using a minimum point and then changing this with random numbers. The random numbers were generated using the built in function `int random()`. As they are equally distributed in $[0, \texttt{RAND\_MAX}]$ we did $\frac{\texttt{value}}{\sqrt{6}} \cdot 2 \cdot \left(\frac{\texttt{random()}}{\texttt{RAND\_MAX}} - 0.5\right)$ with `value=14` as a target distance. The distances to the target point can be found in table 7(a).

Another test series with the same method but the absolute photoconsistency function, a target value of 20 and changes just in the translation components was done. The results can be found in table 7(b).

# phases	t_x	t_y	t_z	r_x	r_y	r_z
1	1.69	2.12	1.57	2.42	2.08	2.22
2	1.66	2.06	1.72	2.33	1.77	2.17
3	1.43	1.96	1.51	1.84	1.9	2.06
4	1.42	1.95	1.65	1.86	1.73	1.91
5	1.4	1.96	1.55	1.97	1.9	2.1
6	1.48	1.93	1.59	1.76	1.82	1.79

(a) phantom renderings

# phases	t_x	t_y	t_z	r_x	r_y	r_z
1	1.98	2.21	2.06	2.07	3.7	2.47
3	1.74	2.05	1.82	1.64	2.9	2.1
5	1.73	1.88	1.82	1.61	2.6	1.96
10	1.64	1.76	1.73	1.48	2.07	1.88

(b) video images

Fig. 5. The normalised total variations for parameter and number of phases for the image sequences.

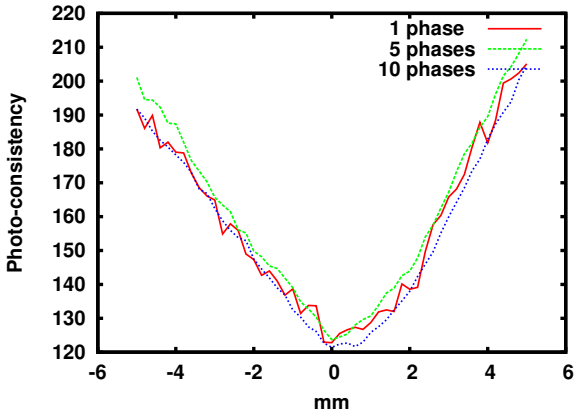


Fig. 6. The image shows translation along the x-axis vs photo-consistency at the minimum point using 1,5, and 10 phases for the video sequence.

before	8.5	4.7	7.8	9.3	9.9
one phase	5.8	2.7	6.4	3.8	8.2
five phases	5.4	2.6	6.4	5.1	7.3

(a) Rotation and translation were changed.

before	8.9	11.6	7.6	2.8	9.3
one phase	5.7	3.9	2.4	2.1	4.6
five phases	4.9	1.0	1.4	1.7	3.5

(b) Just the translation components were changed for the starting value.

Fig. 7. Distances to the minimum points for five sequences.

4 Discussion

From the figures in tables 5(a) and 5(b) we can conclude that the functions are smoother in the meaning of the definition in formula (2). More interesting might be the probable consequences for the convergence which are illustrated in tables 7(a), and 7(b) and show promising behaviour for deviations in the translation component but unfortunately not much improvement if the rotation is included.

An additional complexity in reality is also the heart motion due to breathing. The corresponding movement of the heart can be approximated as rigid. We could therefore hope to be able to separate these two motion components by means of projective geometry.

The confocal light sources for the stereo endoscope of the da Vinci robot are also a significant difficulty, as they are not compatible with our diffuse light model and cause considerable noise. We could simply ignore the brightest parts of the video images as suggested by Stoyanov et al. [7]. Another potential technique would be to calibrate the position of these light sources and to incorporate specular reflections into our lighting model.

Furthermore the endoscopic cameras have a very short baseline of around 5mm. This may make stereo-based registration methods somewhat unstable in the z-direction. This could potentially be overcome using another method, such as external tracking, to estimate the z distance to the surface of the myocardium.

We have presented the notion of using photoconsistency to register intraoperative stereo endoscopic video to a preoperative motion model of the heart surface. Using multiple synchronised frames produces smoother minima in the cost function. We will examine how well this translates into robust 3D registration using both phantom and patient data. This provides a strategy for model-based guidance of TECAB procedures.

References

1. L. Adhami and E. Coste-Maniere. A versatile system for computer integrated mini-invasive robotic surgery. *MICCAI 2002*, 2488:272–281, 2002.
2. J. Bouguet. Camera Calibration Toolbox for Matlab. www.vision.caltech.edu/bouguetj/calib_doc, 2008.
3. M. J. Clarkson, D. Rueckert, D. L. G. Hill, and D. J. Hawkes. Using photo-consistency to register 2d optical images of the human face to a 3D surface model. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23:1266–1280, 2001.
4. J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes. *Computer Graphics*. Addison Wesley, 2nd edition, 1995.
5. J. Jomier, E. Bullitt, M. van Horn, C. Pathak, and S. Aylward. 3D/2D Model-To-Image Registration Applied to TIPS Surgery. In R. Larsen, M. Nielsen, and J. Sporring, editors, *MICCAI 2006*, LNCS (4191), pages 663–9. Springer-Verlag, 2006.
6. F. Mourgues and E. Coste-Maniere. Flexible calibration of actuated stereoscopic endoscope for overlay in robot assisted surgery. *MICCAI 2002, Pt 1*, 2488:25–34, 2002.
7. D. Stoyanov, G. P. Mylonas, F. Deligianni, A. Darzi, and G. Z. Yang. Soft-tissue motion tracking and structure estimation for robotic assisted mis procedures. *MICCAI 2005, Pt 2*, 3750:139–146, 2005.
8. K. Wilson, G. Guiraudon, D. Jones, and T. Peters. 4D Shape Registration for Dynamic Electrophysiological Cardiac Mapping. In R. Larsen, M. Nielsen, and J. Sporring, editors, *MICCAI 2006*, LNCS (4191), pages 520–7. Springer-Verlag, 2006.

A Method for Accelerating Bronchoscope Tracking Based on Image Registration by GPGPU

Takamasa Sugiura[1], Daisuke Deguchi[1], Takayuki Kitasaka[2], Kensaku Mori[1], and Yasuhito Suenaga[1]

[1] Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi, 464-8603, Japan
tsugiura@suenaga.m.is.nagoya-u.ac.jp
[2] Aichi Institute of Technology, Toyota, Japan

Abstract. This paper presents an acceleration method for tracking a bronchoscope based on image registration. This method tracks a bronchoscope by image registration between real bronchoscopic images and virtual ones derived from CT images. However, since the computation cost of image registration, especially generating virtual bronchoscopic (VB) images, is quite expensive, it is difficult to track the bronchoscope in real-time. To solve this problem, we try to accelerate the process of image registration by utilizing GPU (Graphics Processing Unit) with CUDA language. Specifically, we accelerate two parts: (1) VB image generation by volume rendering, and (2) image similarity calculation between a real and a virtual bronchoscopic images. Additionally, to obtain the maximum performance of GPU as much as possible, we minimize (i) the amount of memory transfer between CPU and GPU, and (ii) the number of GPU function calls from CPU. We applied the proposed method to ten pairs of real bronchoscopic video and CT images. The experimental results showed that the proposed method could track the bronchoscope 16 times faster than the method using only the latest CPU.

1 Introduction

In recent years, a bronchoscope navigation system has been developed for assisting a bronchoscopist to operate a bronchoscope safely and efficiently [1–3]. This system gives a bronchoscopist much useful information, such as display of the current location of the bronchoscope's tip, the path to a target location, and anatomical structures beyond bronchi walls that cannot be observed through the bronchoscope.

To realize this system, it is indispensable to track the motion of the bronchoscope camera equipped on the tip of a bronchoscope in real-time. Several research groups reported tracking methods that estimate the motion by the image registration between real bronchoscopic (RB) and virtual bronchoscopic (VB) images [1–3]. Bricault et al. estimated the motion of an RB camera using structural information of bronchial branching patterns [1]. However, their method is not

applicable to locations where branching structures cannot be observed. Also, it cannot estimate camera location for the cases where tumors exist inside airways. Helferty et al. reported a registration method based on similarity measure using normalized mutual information between RB and VB images [2]. Also, Mori et al. proposed an image-registration based method that predicts the camera's motion by the Kalman filter [3]. Prediction results are used for determining initial locations of similarity maximization process to avoid falling into local minima. In the bronchoscope tracking method based on image-registration between RB and VB images, it is necessary to generate considerable number of VB images and to compare them with an RB image for obtaining a good registration result. However, since the computation cost for generating VB images is quite expensive, their methods were far from real-time tracking, for example 8.7 seconds for processing one frame in [3].Therefore, their methods cannot be applied to the bronchoscope navigation in an actual operation. For real-time tracking, we proposed an acceleration method for tracking the bronchoscope based on image registration.

GPU (Graphical Processing Unit), especially GPGPU with special language such as CUDA (NVIDIA Corp.), has great computational capabilities in floating point operation and it has become more than ten times faster than those of CPUs. Therefore, this paper uses computational power of GPU for accelerating the image-registration algorithm.

As described above, image registration process requires to generate huge number of VB images for comparing them with an RB image. There are two methods for generating VB image: (a) surface rendering and (b) volume rendering method. However, when the RB camera is close to the bronchial wall, triangles forming the shape of the bronchus can be seen on VB images in the surface rendering [4]. Due to the artifacts of such triangles that do not appear on volume rendering images, tracking results become less-accurate. Therefore, this paper uses volume rendering in bronchoscope tracking. Since the computation cost for generating a VB image by volume rendering is quite expensive, this generation process is one of the biggest bottlenecks of image registration process. An acceleration of VB image generation can greatly improve the computation time of the image registration process. Therefore, the proposed method tries to accelerate VB image generation process by using CUDA, and uses acceleration technique proposed by Kruger et al.[5]. In the programming on CUDA, GPU is considered a processor that processes a large number of threads simultaneously. Each thread conducts the general-purpose computation by executing a single program (a kernel) of CUDA. In CUDA, there is a memory area where multiple threads can share data. Synchronization of threads on GPU is achieved without any intervention from CPU. Therefore, the data exchange among threads (= shader units) is much easier and cheaper to implement than that of common shader languages such as Cg or HLSL. However, CPU and GPU cannot share memories each other on a conventional PC platform. Data transfer between two types of memories would become a bottleneck of image registration, since memory transfer operations are quite slow than calculation operations on

GPU. Also, there is a small overhead in calling GPU function from CPU. To obtain the maximum performance of GPU as much as possible, we have to minimize (i) an amount of memory transfer between CPU and GPU, and (ii) the number of GPU functions called from CPU. With satisfying these requirements, the proposed method accelerates two parts of image registration: (1) VB image generation by volume rendering, and (2) image similarity calculation between an RB and a VB images by GPU. Since both parts are implemented on GPU, we can directly compute image similarity between RB and VB without giving any controls to CPU. Only a result of image similarity calculation is transferred from GPU to CPU. This greatly saves an amount of memory transfer between CPU and GPU.

2 Method

The GPU has been developed as Graphics Processor which execute graphics process. In graphics process, a lot of data are computed in parallel. Therefore, when GPU computes data weakly depending on each other with high parallelism, we can obtain the maximum performance of GPU. On the other hand, the CPU is good at sequential process which includes complex operations, such as branch instructions. Therefore, when using CPU and GPU, we must consider about CPU's and GPU's advantages. The image registration procedure presented in [4] is simplified as: (1) initialization of view point and view direction, (2) VB image generation, (3) image similarity computation, (4) convergence test and iteration, (5) output of view point and view direction (Fig. 1). Powell method is used for iterative minimization of the image similarity in Step (4) [6]. During volume rendering procedure, each pixels of VB image is calculated independently. Also, many summations during image similarity calculation can be accelerated by parallel computation. Therefore, this paper accelerates the image registration process by implementing the steps (2) and (3) on GPU. Since the steps (2) and (3) are executed in cascade on GPU, we can avoid penalty of data transfer.

2.1 Fast VB image generation by GPU

Volume rendering is one of the techniques for visualizing volumetric images. In prior to rendering, we need to define a transfer function of color and opacity values. Color of each pixel on a screen is calculated by casting a ray from the viewpoint. This ray casting process is performed by accumulating colors and opacities at sample points on the ray.

Volume rendering requires calculations of intensity value and normal vector at each sample point. Linear interpolation technique is usually used for calculating both intensity values and normal vectors. Here, by using linear interpolation technique, an intensity value $f(x, y, z)$ at (x, y, z) can be calculated as

$$f(x, y, z) = \big(1 - (z - \lfloor z \rfloor)\big)\, f(x, y, \lfloor z \rfloor) + (z - \lfloor z \rfloor)\, f(x, y, \lfloor z \rfloor + 1), \qquad (1)$$

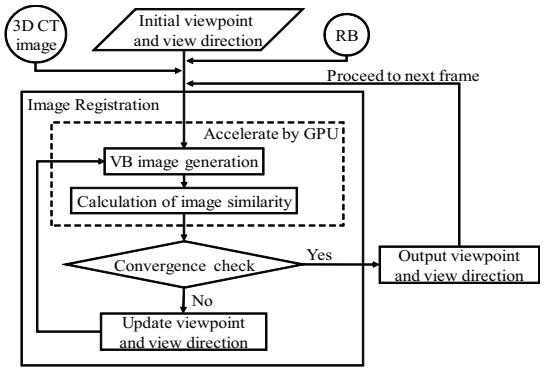


Fig. 1. Processing flow of the proposed method.

where $\lfloor a \rfloor$ is a floor operation, and it returns maximum integer value less than or equal to a. $f(x, y, \lfloor z \rfloor)$ and $f(x, y, \lfloor z \rfloor + 1)$ can be calculated as

$$\begin{aligned} f(x, y, \lfloor z \rfloor) &= w_1 f(\lfloor x \rfloor, \lfloor y \rfloor, \lfloor z \rfloor) + w_2 f(\lfloor x \rfloor + 1, \lfloor y \rfloor, \lfloor z \rfloor) \\ &\quad + w_3 f(\lfloor x \rfloor, \lfloor y \rfloor + 1, \lfloor z \rfloor) + w_4 f(\lfloor x \rfloor + 1, \lfloor y \rfloor + 1, \lfloor z \rfloor), \end{aligned} \tag{2}$$

$$\begin{aligned} f(x, y, \lfloor z \rfloor + 1) &= w_1 f(\lfloor x \rfloor, \lfloor y \rfloor, \lfloor z \rfloor + 1) + w_2 f(\lfloor x \rfloor + 1, \lfloor y \rfloor, \lfloor z \rfloor + 1) \\ &\quad + w_3 f(\lfloor x \rfloor, \lfloor y \rfloor + 1, \lfloor z \rfloor + 1) + w_4 f(\lfloor x \rfloor + 1, \lfloor y \rfloor + 1, \lfloor z \rfloor + 1), \end{aligned} \tag{3}$$

$$\begin{aligned} w_1 &= (1 - (y - \lfloor y \rfloor))(1 - (x - \lfloor x \rfloor)), \\ w_2 &= (1 - (y - \lfloor y \rfloor))(x - \lfloor x \rfloor), \\ w_3 &= (y - \lfloor y \rfloor)(1 - (x - \lfloor x \rfloor)), \\ w_4 &= (y - \lfloor y \rfloor)(x - \lfloor x \rfloor). \end{aligned}$$

As shown in Eqs.(1), (2) and (3), $f(x, y, z)$ is calculated by referring intensity values of surrounding eight voxels. Also, a normal vector of each sampling point is calculated by referring intensity values of surrounding 32 voxels. Although the cost for each linear interpolation is quite small, numerous number of interpolation is required for generating VB images. Therefore, acceleration of linear interpolation will improve the performance of VB image generation. The proposed method tries to accelerate this by linear interpolation hardware of GPU.

To use this linear interpolation hardware that can be accessible from the CUDA language, the proposed method maps a 3D image coordinate $(x, y, \lfloor z \rfloor)$ onto a 2D texture coordinate (x', y') as

$$\begin{aligned} (x', y') &= \phi\,(x, y, \lfloor z \rfloor) \\ &= \left(x + W\Big(\lfloor z \rfloor \bmod N\Big), y + H\lfloor \frac{\lfloor z \rfloor}{N} \rfloor \right), \end{aligned} \tag{4}$$

where N is the number of CT slices placed along the x-axis of 2D texture coordinate system, and W and H are the width and the height of each slice. Figure

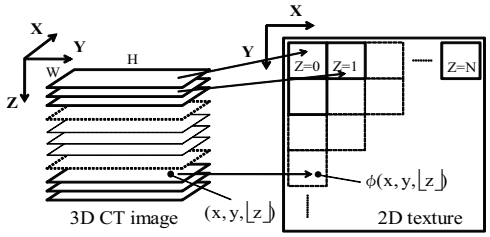


Fig. 2. Relationships between coordinate systems of 3D image and 2D texture.

2 illustrates this mapping function ϕ. As shown in Fig. 2, the proposed method places each slice of the 3D CT image as a 2D matrix. By using this function, we can obtain intensity values directly at arbitrary positions inside the 2D texture coordinate system by GPU hardware. An intensity value $f(x, y, \lfloor z \rfloor)$ can be obtained as

$$f\left(x, y, \lfloor z \rfloor\right) = F\big(\phi\left(x, y, \lfloor z \rfloor\right)\big), \tag{5}$$

where $F\left(\phi\left(x, y, \lfloor z \rfloor\right)\right)$ is an intensity value at a 2D texture coordinate $\phi\left(x, y, \lfloor z \rfloor\right)$. From Eqs.(4) and (5), Eq.(1) can be rewritten as

$$f(x, y, z) = \big(1-(z - \lfloor z \rfloor)\big)\cdot F\big(\phi\left(x, y, \lfloor z \rfloor\right)\big)+(z - \lfloor z \rfloor)\cdot F\big(\phi\left(x, y, \lfloor z \rfloor + 1\right)\big). \tag{6}$$

Since $F\big(\phi\left(x, y, \lfloor z \rfloor\right)\big)$ and $F\big(\phi\left(x, y, \lfloor z \rfloor + 1\right)\big)$ are directly obtained by GPU hardware, calculation of $f(x, y, z)$ is simplified to 1D linear interpolation. Therefore, $f(x, y, z)$ can be calculated by referring only two intensity values at the 2D texture. A normal vector of each sampling point can be calculated by same way.

2.2 Fast image similarity calculation by GPU

The proposed method uses mean-squared error as image similarity measure. Here, image similarity between an RB and a VB image is calculated as

$$MSE\left(\mathbf{B}, \mathbf{V}\right) = \sum_i \left(\left(\mathbf{B}_i - \overline{\mathbf{B}}\right) - \left(\mathbf{V}_i - \overline{\mathbf{V}}\right)\right)^2, \tag{7}$$

where $\mathbf{B}$ and $\mathbf{V}$ are an RB and a VB image. $\mathbf{B}_i$ and $\mathbf{V}_i$ are pixel values at i-th pixel, and $\overline{\mathbf{B}}$ and $\overline{\mathbf{V}}$ are mean pixel value of $\mathbf{B}$ and $\mathbf{V}$, respectively [4]. When we calculate this similarity, it requires two consecutive steps: (a) calculation of $\overline{\mathbf{B}}$ and $\overline{\mathbf{V}}$, and (b) calculation of Eq.(7). That is, two calls of GPU function are required for calculating Eq.(7). To reduce the overhead of GPU function calls from CPU, we rewrite Eq.(7) as

$$MSE\left(\mathbf{B}, \mathbf{V}\right) = \sum_i \Big(\mathbf{B}_i - \mathbf{V}_i\Big)^2 - |\mathbf{B}| \left(\overline{\mathbf{B}}^2 + \overline{\mathbf{V}}^2 - 2\overline{\mathbf{B}}\,\overline{\mathbf{V}}\right), \tag{8}$$

where $|\mathbf{B}|$ is the number of pixels inside $\mathbf{B}$. All terms in Eq.(8) can be calculated independently by GPU, and an image similarity is obtained by summing all

results on CPU. Eq.(8) requires only one GPU function call for calculating image similarity, and we can save an overhead of one GPU function call than Eq.(7). Eq.(8) is computed by the following steps.

[Image similarity calculation]

[**Step** 1] Thread i on GPU calculates

$$a_i = \left(\mathbf{B}_i - \mathbf{V}_i\right)^2, \quad b_i = \mathbf{B}_i, \quad c_i = \mathbf{V}_i.$$

[**Step** 2] Reduction algorithm [7] is used on GPU for summing a_i, b_i, c_i as

$$\alpha = \sum_i a_i, \quad \beta = \sum_i b_i, \quad \gamma = \sum_i c_i.$$

[**Step** 3] α, β and γ are transferred to CPU.

[**Step** 4] Image similarity of Eq.(8) is computed on CPU as

$$MSE\left(\mathbf{B}, \mathbf{V}\right) = \alpha - |\mathbf{B}| \left(\left(\frac{\beta}{|\mathbf{B}|}\right)^2 + \left(\frac{\gamma}{|\mathbf{B}|}\right)^2 - 2\frac{\beta}{|\mathbf{B}|}\frac{\gamma}{|\mathbf{B}|} \right).$$

3 Results and Discussion

We applied the proposed method to ten pairs of RB videos and CT images of the same patient. The specifications of CT images were: 512×512 pixelsC72 $\sim$ 361 slicesC2.0 $\sim$ 5.0 mm slice thicknessCand 1.0 $\sim$ 2.0 mm reconstruction pitchD RB videos were recorded onto digital video tapes and captured at 30 frames per second. All registration tasks were performed as off-line jobs. We used a Dell Precision workstation (Intel Quad Core Xeon 2.80 GHz×2CGeForce 8800 GTXC3.0 GB RAM, Windows XP), and implemented the proposed algorithm as GPU codes by using CUDA.

To evaluate the computation efficiency of the proposed method, we measured the average computation time for generating VB images by changing image size and observation parameters, such as camera position and orientation. Figure 4 shows the computation time for generating VB images on CPU and GPU. Also, we measured the computation time of the image registration process of the previous and the proposed methods. The results of both methods are shown in Table 1. Figure 3 shows the tracking results obtained by the previous and the proposed methods.

Figure 4 shows that the bigger image size is, the more remarkable the difference between computation time for VB image generation on CPU and GPU is. Also, VB image (128×128) generation on GPU was performed 22 times faster than on CPU. From this result, it is easily confirmed that GPU enabled us to generate VB images quite faster than CPU. As shown in Table 1 and Fig. 3, although the computational speed of the proposed method was quite faster than that of the previous method, we obtained almost the same tracking results from

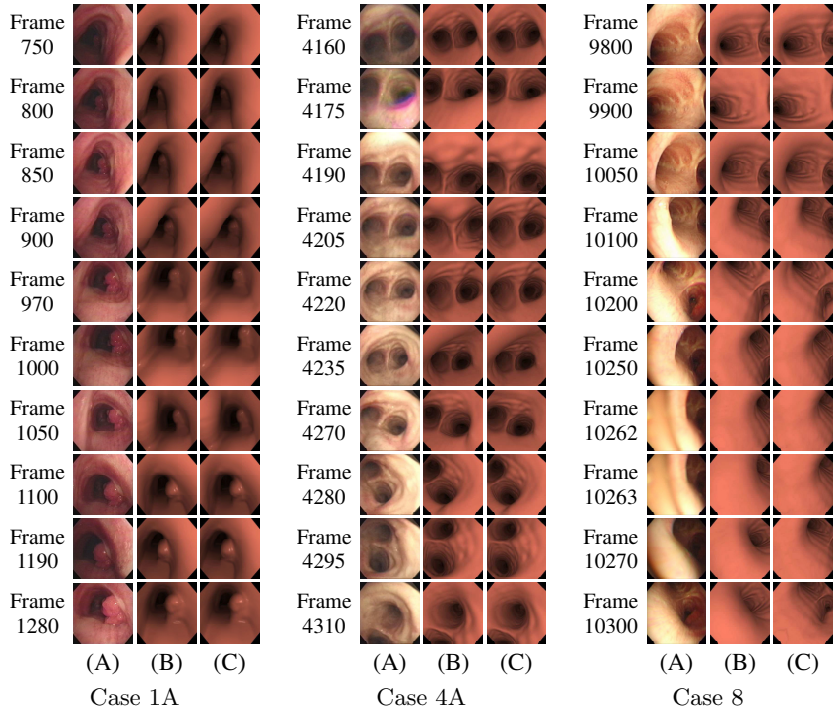


Fig. 3. Tracking results. (A) RB image, (B) results by the previous method, and (C) results by the proposed method.

the CPU-base and the GPU-base methods. However, there are small differences among the number of success frames tracked continuously in these methods. Since the accuracy of the linear interpolation hardware of GPU is not as accurate as CPU does (effective digits), VB images generated by GPU were slightly different from those of CPU. This caused different results despite of doing same procedures.

From Table 1, the average computation time was 0.068 sec. in the proposed method, while 1.085 sec. in the previous method. This means that the proposed method is 16 times faster than the previous method. From these results, we can confirm that GPU is quite helpful for accelerating bronchoscope tracking based on image registration. However, the proposed method may be more accelerated by modifying the process of the image registration for GPU. For example, by changing optimization problem to parallel algorithm, such as [8], we may accelerate the process on GPU.

In general, the tracking speed should be faster than video rate (30 frames/sec.) for real-time guidance. However, the proposed method could only track the bronchoscope at 15 frames per second. We compare the tracking performances of two methods: (a) execution of image registration for every frame and (b) for every

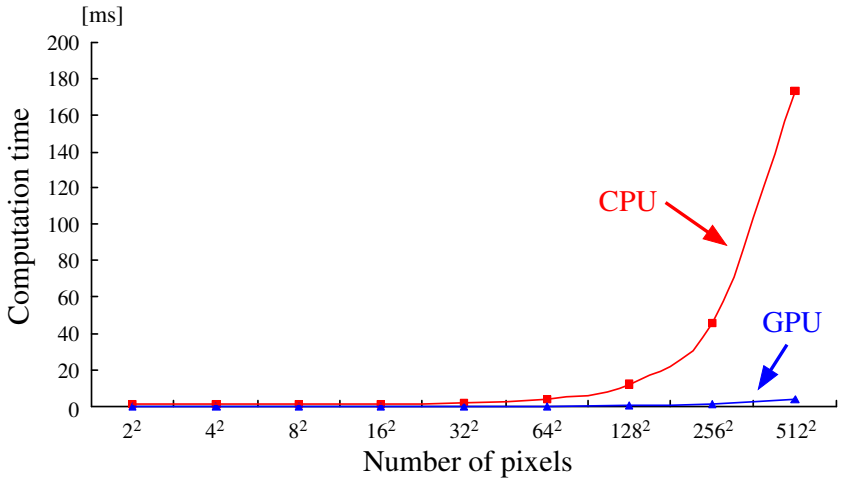


Fig. 4. The computation time for generating VB images when image size changes.

two frames. The results showed that the method (b) could track frames is 5724 successive frames in total. Although the tracking performance decreased to 89 % of the method (a), this results demonstrate the possibility of the method (b) for real-time tracking. On the other hand, since this paper uses single initial guess for image registration, tracking performance of the proposed method was poor than that of [3]. To improve tracking performance, we should try to implement powerful image similarity on GPU. This is also our future work.

4 Conclusions

This paper presented a method for accelerating bronchoscope tracking based on image registration by GPU. To accelerate the process of bronchoscope tracking, we implemented the bottleneck procedures of image registration on GPU, such as generation of VB images by volume rendering and image similarity calculation. In the proposed method, volume rendering can be accelerated efficiently by fast linear interpolation of GPU hardware, and the process of bronchoscope tracking is also accelerated by the reduction of the traffic between CPU and GPU. The experimental results showed that the computational speed of the proposed method was 15 frames per second on average.

Future work includes: (a) investigation of task assignment between CPU and GPU, (b) accuracy improvement of the image registration process.

References

1. I. Bricault, G. Ferretti and P. Cinquin, “Registration of Real and CT-Derived Virtual Bronchoscopic Images to Assist Transbronchial Biopsy,” IEEE Transactions on

Table 1. Results of bronchoscope tracking. The numbers show consecutive frames successfully (average computation time [sec.]). The number of successive frames were counted with our visual inspection.

Case	Path	Num. of frames	Results		Improvement ratio
			Previous	Proposed	
1	A	993	575 (0.795)	572 (0.078)	0.99 (10.2)
	B	700	154 (1.223)	154 (0.081)	1.00 (15.1)
	C	1120	45 (1.586)	24 (0.077)	0.53 (20.6)
	D	379	70 (0.892)	70 (0.057)	1.00 (15.6)
	E	279	93 (0.828)	94 (0.079)	1.01 (10.5)
2	A	429	14 (1.318)	14 (0.058)	1.00 (22.7)
3	A	1500	119 (1.327)	119 (0.064)	1.00 (20.7)
	B	1100	359 (1.085)	361 (0.050)	1.01 (21.7)
	C	1000	279 (1.203)	136 (0.050)	0.49 (24.1)
	D	202	32 (1.850)	32 (0.071)	1.00 (26.1)
	E	250	111 (1.607)	111 (0.057)	1.00 (28.2)
4	A	840	161 (1.518)	161 (0.108)	1.00 (14.1)
	B	204	204 (1.044)	204 (0.067)	1.00 (15.6)
5	A	500	22 (1.748)	22 (0.085)	1.00 (20.1)
6	A	1000	1000 (1.062)	718 (0.084)	0.72 (12.6)
	B	800	44 (1.034)	44 (0.061)	1.00 (17.0)
	C	1000	13 (0.941)	41 (0.075)	3.15 (12.5)
7	A	500	89 (1.151)	89 (0.083)	1.00 (13.9)
	B	530	140 (1.119)	140 (0.071)	1.00 (15.8)
8	A	1600	462 (0.621)	536 (0.056)	1.16 (11.1)
9	A	450	125 (0.870)	125 (0.061)	1.00 (14.3)
10	A	900	194 (0.690)	315 (0.074)	1.62 (9.3)
	B	1400	558 (0.433)	695 (0.055)	1.25 (7.9)
	C	2000	1183 (0.578)	1182 (0.044)	1.00 (13.1)
	D	820	499 (0.611)	499 (0.047)	1.00 (13.0)
Total (ave.)			6545 (1.085)	6458 (0.068)	0.99 (16.2)

Medical Imaging, Vol. 17, No. 5, pp. 703-714, 1998.
2. J. P. Helferty and W. E. Higgins, "Technique for Registering 3D Virtual CT Images to Endoscopic Video," Proceedings of ICIP (International Conference on Image Processing), pp. 893-896, 2001.
3. K. Mori, D. Deguchi, T. Kitasaka et al, "Bronchoscope Tracking Based on Image Registration Using Multiple Initial Starting Points Estimated by Motion Prediction," Proc. of MICCAI 2006, LNCS, Vol. 4191, pp. 645-652, 2006.
4. D. Deguchi, K. Mori, J. Hasegawa et al, "Camera motion tracking of real bronchoscope using epipolar geometry analysis and CT derived bronchoscopic images," Proceeding of SPIE Vol. 4683, pp. 30-41, 2002.
5. J.Kruger and R. Westermann, "Acceleration Techniques for GPU-based Volume Rendering," IEEE Visualization, pp. 287-292. 2003.
6. W. H. Press, S. A. Teukolsky, W. T. Vetterling et al, "Numerical Recipes in C," The Art of Scientific Computing Second Edition, CAMBRIDGE UNIVERSITY PRESS, pp. 321-336, 1999.

7. Mark Harris, "Optimizing Parallel Reduction in CUDA," CUDA Advanced Topics, CUDA ZONE,
"http://www.nvidia.com/object/cuda_sample_advanced_topics.html".
8. F. V. Berghen, H. Bersini, "CONDOR, a new parallel, constrained extension of Powell's UOBYQA algorithm: Experimental results and comparison with the DFO algorithm," Journal of Computational and Applied Mathematics, Elsevier, Volume 181, Issue 1, September 2005, pp. 157-175.

Semi-Automated Segmentation of Pelvic Organs for AR Guided Robotic Prostatectomy

Wenzhe Shi*, Lin Mei*, Ara Darzi, and Philip Edwards*

Dept. of Biosurgery and Surgical Technology Imperial College London, UK
ws207,l.mei,eddie.edwards@imperial.ac.uk

Abstract. Building subject-specific 3D models for pelvic organs from preoperative imaging is a vital step in achieving augmented reality guidance for robotic prostatectomy. This paper presents a semi-automatic method for calculating deformations from atlas models to subject specific models for pelvis, lower spine, coccyx, rectum and prostate. The method consists of three stages and requires no more than 5 minutes human intervention in identifying 31 landmarks located in the lower abdominal region. The first step is to perform point-based affine and non-rigid registration to produce an initial deformation for the whole pelvic region. The second step refine this deformation with two levels of intensity-based non-rigid registration, the result of which is then used as the input to third stage where organ-specific deformations are produced by non-rigid registrations applied to the region of interest. We applied this method to build subject-specific models for 18 patients. When compared with manual segmentation produced by expert clinicians, the result showed around 20% improvement over the baseline method, which uses only intensity-based affine and non-rigid registration, and more than more than 90% on some case where the baseline method failed. We believe this is a significant step towards clinically usable segmentation of lower abdominal organs for augmented reality guided robotic prostatectomy.

1 Introduction

Prostate cancer is an increasing problem in an aging society. Amongst the possible treatments, minimally invasive robotic radical prostatectomy is rapidly gaining acceptance [1]. There are, however, significant rates of incontinence, impotence or incomplete resection [2]. It is crucial that the surgeon accuratley targets the prostate, particularly near the prostatic apex, and that the surrounding neuro-vascular bundle is well preserved.

These difficulties can be eased using intraoperative image guidance. It is suggested that augmented reality (AR) provides better guidance though improved visualisation [3]. AR technology overlays the 3D models of important structures onto surgeon's field of view. The fact that Da Vinci system uses a real-time stereo video captured by endoscope as surgeon's view makes it a natural platform for implementing AR guidance.

One major step towards AR surgical guidance is to build patient-specific 3D models of the important organs located in the operational region. These

models are normally built by segmentation of preoperative images such as MRI and CT. Segmentation is a huge research field. One option is the use of an atlas. A template segmentation can be registered to an individual, which has been applied to segmentation of brain [4] and lung tissues[5]. Segmentation of kidneys has been proposed using a a gradient-fitting deformable atlas [6], and also segmentation of other abdominal organs such as liver, kidneys, and spinal cord using a thing-plate spine based probabilistic atlas [7].

Research in segmentation of the prostate and surrounding pelvic organs is driven mainly by demand in radiotherapy. Intensity-based methods are proposed for MRI [8, 9]. A statistical atlas has also been proposed. Shen et. al. built a statistical shape model from ultrasound images [10]. Arambula proposed an active shape model that is optimised by a genetic algorithm [11]. Pasquier et. al. compared a seed growing method to an adaptive atlas based on morphological filtering for the prostate and other abdominal organs on MRI. They found atlas-based segmentation to be more robust but it requires minor interactive corrections, while region-growing only works fine for the bladder and rectum, for which MRI shows good soft-tissue contrast [12]. Their model is further extended to a probabilistic atlas which deforms only along a certain principle modes in the training set [13].

To calculate the deformation that warps an atlas to a specific image, one can use a direct search along the degrees of freedom(DOI) of the atlas. This method allows unrealistic deformation to be introduced due to unwanted local maxima. A possible enhancement is to train a statistical atlas which only deforms within a feasible shape space. However, this enhancement requires a significant training set. Otherwise restricted deformation may result in failed segmentation for uncommon shapes in the presence of pathology.

In this paper we present a semi-automatic segmentation-by-registration scheme for building models of pelvic organs including lower spine, coccyx, rectum including rectal-sigmoid junction, and prostate. The original atlas is built on visible human (VH) [14], and the targeted images are preoperative MR scans of 18 patients. Despite varying resolution, different MR sequences and pathological shape variation, our method still manged to produce feasible segmentation after minor manual corrections.

2 Method

Segmentation by registration has been proposed for the lung [5]. Since this method does not restrict atlas deformation, it can be used for segmenting pathological images. A hierarchy of affine and non-rigid registration proposed by Rueckert et. al. [15] is used to calculate the transformation that warps the atlas 3D model to fit the subject's anatomy.

The non-rigid transformation is described by a regular grid of 3D control points. The B-spline interpolates the displacements inbetween the control points. The displacement of a control point only affects its local neighbourhood. Hence a single b-spline transformation can represent both local and global deformation.

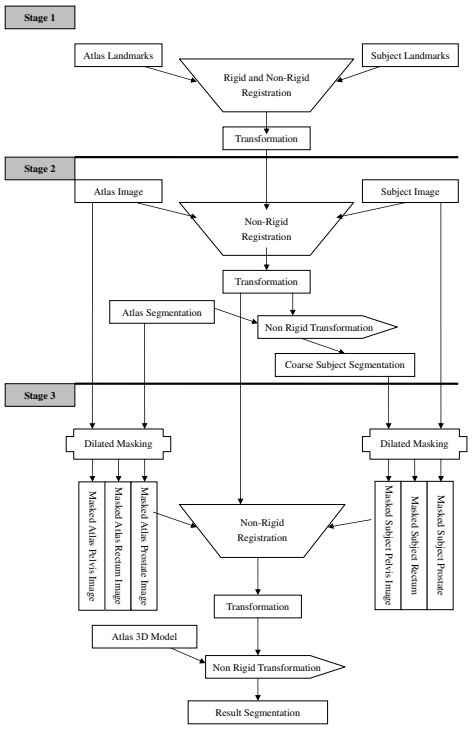


Fig. 1. Flow chart of the registration process

Also, since the atlas remains the same, the control points of transformation for different individuals naturally correspond. Such correspondence is a requirement for statistical shape analysis.

In order to produce an accurate transformation, the hierarchical registration consists of three stages - landmark based registration, intensity-based coarse registration and a series of registrations performed for each organ (see Fig. 1). Note that the direction of registration is in fact from subject to atlas because the resulting transformation calculates the corresponding points in the source image space for a target image voxel.

2.1 Stage I: Landmark-based Registration

Corresponding landmarks are manually selected in the atlas and the subject images. This provides an initial registration. A total of 31 landmarks are used: 18 on the pelvis, 2 on either end of the rectum and 11 on the prostate. The selection of the landmarks on the pelvis and rectum avoids common misregistrations

resulting from the intensity based approach. The selection of the landmarks on prostate is to cope with some extremely pathological shape variation.

Affine registration is performed on the landmarks to aligned the position and orientation. This is followed by a non-rigid registration using 20mm grid size of which the output transformation is applied to the subject image to produce a starting point for the next stage. We will describe the reason for choosing this grid size in the following section.

2.2 Stage II: Intensity-Based Hierarchical Non-Rigid Registration

The second stage is to perform a series of non-rigid registrations between the atlas and subject's MR scans. T2-weighted MRI is used because it is one of the most commonly used diagnostic images and has better soft tissue contrast than CT or T1-weighted MRI and because the central glad and peripheral zone are distinguishable [16]. This makes it possible to produce a detailed segmentation of the prostate.

The registration uses a B-spline as the non-rigid transformation and normalised mutual information (NMI) as the similarity measure. NMI is selected since it is proven to be the most effective similarity measures for registration of between images with different modalities [17, 18].

After applying the resulting transformation of stage I onto the subject's image, an intensity-based non-rigid registration is performed.

Since displacing the control points would only affect their neighbourhoods and the smoothness of the local deformation is affected by control point spacing, larger spacing leads to a smoother and more global deformation while a relatively smaller spacing leads to a more localised but less smooth deformation. The nature of the gradient descent optimisation approach with NMI gives rise to slow convergence properties and increased sensitivity to local minima. These unattractive features are especially evident when trying to align large and complex images. One standard enhancement is a hierarchical approach.

The basic idea is to start with a coarse initial grid size and sub-sampled image, and refine the transformation iteratively with increasing image resolution and decreasing grid size. In our experiments we increase image resolution and reduce the grid size both by a factor of two at each iteration. A large grid size is required to avoid a highly localised deformation. However, when grid size is too large, there will be more than one landmark in one grid cell and the b-spline function fails to interpolate their displacement. It was found that when grid size is too large, the result of stage I may not be sufficiently accurate.

Taking into account the trade-off between avoiding a very localised deformation and providing interpolation of the dense landmarks on the prostate, the settings in table 1 are used:

2.3 Stage III: Registration on the Region of Interest

A mask of the atlas is created for each organ by dilating the binary segmentation. This provides a greyscale image that includes the organ and some surrounding

	Coarse Level	Refined Level
Stage I	20mm grid size with 5mm voxel size	N/A
Stage II	20mm grid size with 5mm voxel size	10mm grid size with 2.5mm voxel size
Stage III	10mm grid size with 2.5mm voxel size	5mm grid size with 1.25mm voxel size

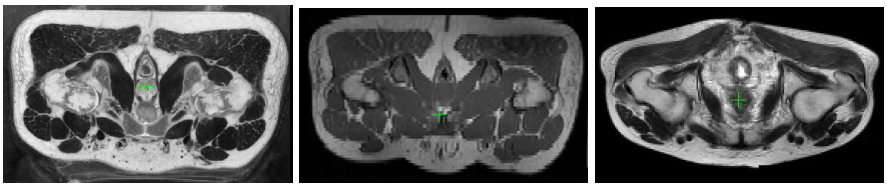
Table 1. Grid size with corresponding voxel size

tissue. Separate images are created for the pelvis (including partial spine and coccyx), the rectum, and the prostate. This reduces the unwanted influence of tissues more distant from the target organs. This also speeds up the registration process, allowing smaller grid size and voxel dimension to be used.

In the last stage, we apply the resulting transformation from stage II to the subject's image and register it onto each masked atlas respectively to get the transformation for each organ. Applying these transformations to the 3D model built from atlas segmentation we get the subject specific model.

3 Experiment and Results

3.1 Atlas Construction and Subject Image Preprocessing



A B C

Fig. 2. Visible Human images: Cryo-section (A), T-1 MRI (B) and a combined MRI of a subject(C).

We used the Visible Human (VH) cryo-sectional image for manual segmentation because it provides good visualisations of the pelvic soft tissues (see figure 2). The pelvis is segmented and registration performd using the VH T1 MRI because of poor soft-tissue contrast in CT and poor image quality for T2 MRI.

We acquired lower abdomen MRI scans for 18 patients. There are multiple scans per subject, generally 5mm axial and sagittal slices with high in-plane resolution ($0.8mm \times 0.8mm$) together with a few focused T2 scans of the prostate providing higher resolution. In order to take advantage of all these images, we resample to the highest resolution and calculate a weighted average image that is used for registration. An example is shown in figure 2 C.

The segmentation procedure described in section 2 is used to propagate the manual segmentation to all the subjects. Since the modality of VH MRI and the patients' MRIs are very different, we chose the patient with be best registration

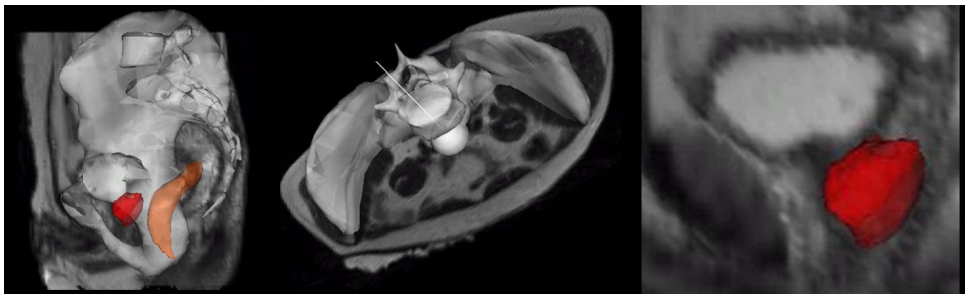
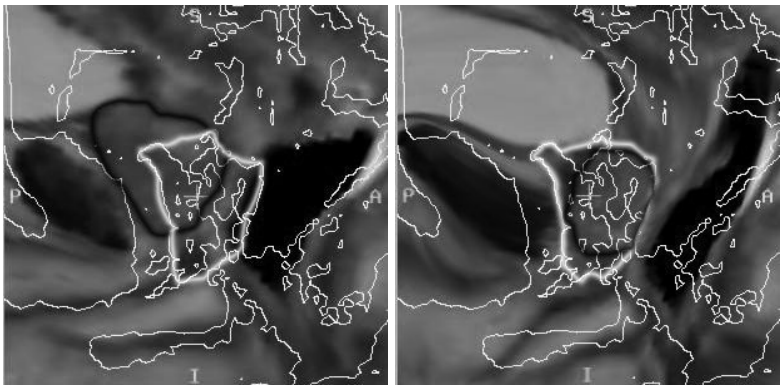


Fig. 3. An example of the segmented subject model.



A B

Fig. 4. Segmentation of prostates (marked with bright rings) using direct intensity-based registration (A) and using landmark(B) before intensity-based registration. The true prostate positions are marked by dark rings

result to use as the atlas and propagate this segmentation result to its peers. Results are shown below in section 3.2.

3.2 Segmentation by Registration

Figure 3 shows an example of patient specific model. We compared the result of our method with a baseline method using the direct intensity-based affine registration followed by two levels of non-rigid registration.

Figure 4 shows results with and without landmark guidance are compared, and figure 5, shows results with and without a masked atlas are compared. Dramatic improvement in accuracy are found in both.

Quantitative evaluation is done by comparing the results to manual segmentation produced for 4 subjects by clinicians. Using normalised partition distance [19] as measure, our method showed around 35% improvement over the baseline, and significant improvement (more than 90%) on some case where simple intensity-based registration failed(see table 2).

A B

Fig. 5. Segmentation of prostates (marked with bright rings) without using masked atlas (A) and using masked atlas(B).

	Patient 003	Patient 005	Patient 007	Patient 008
Baseline	9.0%	49.1%	3%	2.2%
Stage I	10.4%	10.2%	6.4%	3.5%
Stage II	7.3%	7.4%	5.9%	2.5%
Stage III	4.9%	4.2%	4.5%	1.8%

Table 2. Example partition distances (in voxels) between the manual gold standard and the semi-automatic segmentation after each registration step.

4 Discussion

We proposed a useful hierarchical segmentation-by-registration scheme for pelvic organs. Compared to manual segmentation, improvement is found at each stage. Possible further developments include automatic landmark identification and construction of a statistical shape or deformation model to restrict the degrees of freedom of the transformation.

References

1. El-Hakim, A., Tewari, A.: Robotic prostatectomy – a review. MedGenMed **6** (2004) 20
2. Kaul, S., Savera, A., Badani, K., Fumo, M., Bhandari, A., Menon, M.: Functional outcomes and oncological efficacy of vattikuti institute prostatectomy with veil of aphrodite nerve-sparing: an analysis of 154 consecutive patients. BJU Int. **97** (2006) 467–472
3. Nowinski, W.: Virtual reality in brain intervention: models and applications. Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the (Jan. 2005) 4164–4167

4. Cuadra, M., Pollo, C., Bardera, A., Cuisenaire, O., Villemure, J.G., Thiran, J.P.: Atlas-based segmentation of pathological mr brain images using a model of lesion growth. Medical Imaging, IEEE Transactions on **23**(10) (Oct. 2004) 1301–1314
5. Sluimer, I., Prokop, M., vanGinneken, B.: Toward automated segmentation of the pathological lung in ct. Medical Imaging, IEEE Transactions on **24**(8) (Aug. 2005) 1025–1038
6. Gao, L.M., Heath, D.G., Fishman, E.K.: Abdominal image segmentation using three-dimensional deformable models. Invest. Radiol. **33** (1998) 348–355
7. Park, H., Bland, P., Meyer, C.: Construction of an abdominal probabilistic atlas and its application in segmentation. Medical Imaging, IEEE Transactions on **22**(4) (April 2003) 483–492
8. Chiu, B., Freeman, G., Salama, M., Fenster, A., Rizkalla, K., Downey, D.: A segmentation algorithm using dyadic wavelet transform and discrete dynamic contour. Electrical and Computer Engineering, 2003. IEEE CCECE 2003. Canadian Conference on **3** (May 2003) 1481–1484
9. Samiee, M., Thomas, G., Fazel-Rezai, R.: Semi-automatic prostate segmentation of mr images based on flow orientation. Signal Processing and Information Technology, 2006 IEEE International Symposium on (Aug 2006) 203–207
10. Shen, D., Zhan, Y., Davatzikos, C.: Segmentation of prostate boundaries from ultrasound images using statistical shape model. Medical Imaging, IEEE Transactions on **22**(4) (April 2003) 539–551
11. Cosío, F.A. In: Prostate Segmentation Using Pixel Classification and Genetic Algorithms. Springer Berlin / Heidelberg (2005) 910–917
12. Pasquier, D., Lacornerie, T., Vermandel, M., Rousseau, J., Lartigau, E., Betrouni, N.: Segmentation of prostate boundaries from ultrasound images using statistical shape model. International Journal of Radiation Oncology Biology Physics **68**(2) (April 2007) 592–600
13. Betrouni, N., Puech, P., Dewalle, A., Lopes, R., Dubois, P., Vermandel, M.: 3d automatic segmentation and reconstruction of prostate on mr images. Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE (Aug. 2007) 5259–5262
14. Spitzer, V., Ackerman, M., Scherzinger, A., Whitlock, D.: The visible human male: atechnical report. Journal of the American Medical Informatics Association **3**(2) (1996) 118–30
15. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J.: Nonrigid registration using free-form deformations: Application to breast mr images. IEEE Trans. Med. Imag. **18**(18) (Aug 1996) 712–721
16. Mattson, J., Simon, M.: The Pioneers of NMR and Magnetic Resonance in Medicine: The Story of MRI. Jericho & New York: Bar-Ilan University Press (1996)
17. Penney, G., Weese, J., Little, J., Desmedt, P., Hill, D., hawkes, D.: A comparison of similarity measures for use in 2-d-3-d medical image registration. Medical Imaging, IEEE Transactions on **17**(4) (Aug 1998) 586–595
18. Zitova, B., Flusser, J.: Image registration methods: a survey. Image and Vision Computing **21**(11) (October 2003) 977–1000
19. Cardoso, J., Corte-Real, L.: Toward a generic evaluation of image segmentation. Image Processing, IEEE Transactions on **14**(11) (Nov. 2005) 1773–1782

Author Index