Proceedings of

AMI-ARCS

September 24, 2009 in London

5th Workshop on Augmented Environments for Medical Imaging including Augmented Reality in Computer-Aided Surgery,

Imperial College London

CAMP

MoRiG

Fraunhofer MEVIS

INITION
EVERYTHING IN 3D
www.inition.co.uk

MICCAI

Program Committee

Adrien Bartoli,
LASMEA, Clermont-Ferrand

Dirk Bartz,
University of Leipzig

Marie-Odile Berger,
LORIA

Wolfgang Birkfellner,
Medical University Vienna

Gabor Fichtinger,
Queens University, Kingston

Markus Hadwiger,
VRVis Research Center

Sandro Michael Heining
LMU, München

Pierre Jannin,
IRISA

Ron Kikinis,
BWH/Harvard

Gordon Kindlman,
University of Chicago

Mirna Lerotic,
Imperial College, London

Stephane Nicolau,
IRCAD

Nassir Navab,
CAMP, Technische Universität, München

Bernhard Preim,
University of Magdeburg

Felix Ritter,
Fraunhofer MEVIS

Jannick Rolland,
University of Rochester

Yoshinobu Sato,
University of Osaka

Anna Vilanova,
Eindhoven University of Technology

Kirby Vosburgh,
CIMIT

General Chair

PJ “Eddie” Edwards
Imperial College London, UK

Kensaku Mori
Nagoya University, Japan

Christoph Bichlmeier
TUM, Germany

Jan Klein,
Fraunhofer MEVIS, Bremen, Germany

http://campwww.informatik.tu-muenchen.de/AMIARCS09/

Preface

These proceedings cover the AMI/ARCS one-day satellite workshop of MICCAI 2009. This workshop continues the tradition of Augmented Reality in Computer Aided Surgery (ARCS) 2003 and the Workshops on Augmented environments for Medical Imaging and Computer-aided Surgery (AMI-ARCS) 2004, 2006 and 2008. 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 physicians view, establishing a direct relation between the image and the patient. The workshop was divided into four sessions - Applications; Efficiency in AR and Registration; Visualisation and Ultrasound; and Visualisation and Algorithms.

There were three invited speakers. Professor Henry Fuchs from UNC Chapel Hill is one of the pioneers of augmented reality research and provided a highly entertaining look at augmented reality research, both past and present. Professor Eduard Gröller of the Technical University Vienna gave a talk on multimodal image fusion for diagnosis of coronary artery disease. This reflects a growing interest amongst the AMI-ARCS community in multimodal image fusion, which was specifically highlighted as a theme this year. Finally, Dr. Hongen Liao of the University of Tokyo described his significant research experience in augmented reality under the heading of integral videography.

A very popular feature of AMI-ARCS 2009 was the inclusion of a demonstration session over lunch. This enabled researchers to experience five cutting edge systems first hand. It is hoped that this will be a feature of future AMI-ARCS workshops. There was industrial involvement from Inition.co.uk who provided 3D visualisation for the workshop. It is hoped that this relationship may continue. At the end of the meeting there was a panel discussion on the future of both AMI-ARCS itself and AR.

AMI/ARCS 2009 brought together clinicians and technical researchers from industry and academia with interests in computer science, electrical engineering, physics, and clinical medicine to present state-of-the-art developments in this ever-growing research area. There is significant ongoing interest in AMI/ARCS, with a steady number of participants and the continued high quality of submissions. We are also pleased with the trend in the research to provide fully functional demonstrations and aiming towards real clinical applications that will be of benefit to patients.

September 24th 2009

PJ "Eddie" Edwards
Christoph Bichlmeier
Jan Klein
Kensaku Mori
Program Chairs
AMI-ARCS 2009

Program Committee

Adrien Bartoli, LASMEA, Clermont-Ferrand
Dirk Bartz, University of Leipzig
Marie-Odile Berger, LORIA
Wolfgang Birkfellner, Medical University Vienna
Gabor Fichtinger, Queens University, Kingston
Markus Hadwiger, VRVis Research Center
Sandro-Michael Heining, LMU, München
Pierre Jannin, IRISA
Ron Kikinis, BWH/Harvard
Gordon Kindlman, University of Chicago
Mirna Lerotic, Imperial College, London
Stephane Nicolau, IRCAD
Nassir Navab,CAMP, Technische Universitt, München
Bernhard Preim, University of Magdeburg
Felix Ritter, Fraunhofer MEVIS
Jannick Rolland, University of Rochester
Yoshinobu Sato, University of Osaka
Anna Vilanova, Eindhoven University of Technology
Kirby Vosburgh,CIMIT

Sponsoring Institutions

Imperial College, London, UK
TUM, Germany
Fraunhofer MEVIS, Bremen, Germany
Nagoya University, Japan
Inition.co.uk London, UK

Table of Contents

Session 1 – Applications

Overview of AR-Assisted Navigation: Applications in Anesthesia and Cardiac Therapy

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

Imaging Research Laboratories, Robarts Research Institute,
Biomedical Engineering Graduate Program, University of Western Ontario,
London ON Canada
{jmoore,tpeters}@imaging.robarts.ca

Abstract. The global focus on minimizing invasiveness associated with conventional treatment has inevitably led to restricted access to the target tissues. Complex procedures that previously allowed surgeons to directly see and access targets within the human body must now be "seen" and "reached" via medical imaging and specially designed tools, respectively. A successful transition to less invasive therapy is subject to the ability to provide physicians with sufficient navigation information that allows them to maintain therapy efficacy despite "synthetic" visualization and target manipulation. Here we describe a comprehensive system currently under development that integrates the principles of augmented reality-assisted navigation into clinical practice. To illustrate its direct benefits in the clinic, we describe the implementation of the system for an anesthetic delivery application, as well as for intracardiac therapy.

1 Introduction

In many engineering applications, computer generated displays are used to model mechanical designs, simulate interaction between different components, and virtually analyze a system's behaviour prior to its implementation; however, these models are rarely carried further into the physical implementation of the system. Similarly, in the medical world, pre-operative information is acquired in order to diagnose a patient's condition, but it is often limited to the treatment plan rather than being integrated into the therapy delivery stage. Augmented reality (AR) technology permits the supply of additional information to the visual field of a user in order to facilitate, or in some cases enable task performance [1]. Using such environments, pre-operative anatomical models featuring the entire treatment plan can be "brought in" during therapy, facilitating treatment or providing physicians with otherwise unavailable information.

Here we build upon the typical AR concepts towards more complex environments. As opposed to using a direct view of the anatomy as "reality", we employ ultrasound (US) imaging to obtain a real-time anatomical display, and further augment the echo images with pre-operative anatomical information and representations of surgical tools. The generated environment not only complements the real-time US display with pre-operatively planned information, but also allows the physician to immerse him/herself within a "mixed reality" environment accurately registered to the patient.

We have demonstrated the implementation of this technology for intracardiac guidance [2], and have already identified the potential that such AR-assisted US guidance environments may eliminate the need for intra-operative fluoroscopic imaging and allow the fusion of surgical planning and guidance [3, 4]. In this paper we review the components of the AR-assisted platform and, in addition to the cardiac application, we also present its implementation for an anesthetic delivery application: facet joint injections and peripheral nerve block.

2 Surgical Platform Architecture And Applications

Our surgical guidance platform (*AtamaiViewer* - http://www.atamai.com) comprises a user interface based on Python and the Visualization Toolkit (VTK) and integrates a wide variety components for IGS applications, including multi-modality image visualization, anatomical modeling, surgical tracking, and haptic control. The environment supports stereoscopic visualization of volumetric data via cine sequences synchronized with the intra-operative ECG [5], as well as fusion of multiple components with different translucency levels for overlays.

The origins of our augmented reality surgical platform arise from the desire of our surgery colleagues to develop a procedure that enables therapy delivery inside the beating heart. Not only does this attempt require a means of introducing multiple tools into the cardiac chambers, but also the ability to visualize and manipulate these devices in real time. Consequently, we adapted these original techniques for other applications, including electro-physiology mapping for atrial fibrillation therapy [3] and an AR system for port placement [6].

2.1 Surgical Guidance: Intracardiac Therapy Delivery

Our first and most challenging AR-assisted application is the planning and guidance of minimally invasive, off-pump mitral valve (MV) implantation and atrial septal defect (ASD) repair. 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.

We proposed an intracardiac visualization approach that relied on intra-operative echocardiography for real-time imaging, augmented with representations of the surgical tools tracked in real-time and displayed within anatomical context available from pre-operative images. By carefully integrating all components, such an environment is capable of expected to provide reliable tool-to-target navigation, followed by accurate on-target positioning.

Using Imaging to "See" Ultrasound is widely employed as a standard interventional imaging modality and 2D TEE (trans-esophageal echocardiography) provides good-quality images, while eliminating the interference between probe manipulation and surgical work-flow. However, the main drawback is its inability to depict sufficiently crisp representations of the anatomical targets and surgical tools, which are highly amplified *in vivo* by the complexity and variability of the anatomy, image quality, and orientation of the US beam with respect to the anatomy. To overcome these problems, 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.

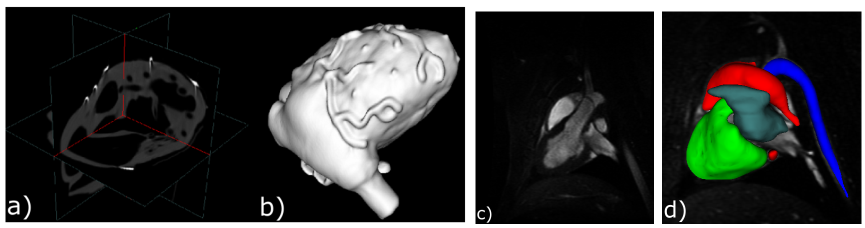


Fig. 1. a) CT image of beating heart phantom; b) Surface model of phantom extracted from CT image; c) Pre-operative cardiac MR image of porcine subject at mid-diastole; d) Extracted cardiac model showing cardiac chambers of interest. *Figure adapted from Linte et al. MICCAI 2008 and Linte et al. CARS 2008.*

For *in vitro* phantom studies, we rely primarily on pre-operative CT images for model building, given their good image contrast and temporal resolution, and make use of automatic techniques (Vascular Modeling Toolkit[1]) for image segmentation. On the other hand, for *in vivo* clinical porcine studies, the excellent soft tissue characterization capabilities of MRI facilitate 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 (**Fig. 1**). Typically, the features of interest include the left ventricle (LV), left atrium (LA), and right atrium and ventricle (RA/RV).

Fusing Pre- and Intra-operative Information We employ a peri-operative feature-based registration technique to augment the intra-operative US images (**Fig. 2a**) with the pre-operative cardiac models (**Fig. 2b**). Easily identifiable targets in both datasets, the mitral (MVA) and aortic (AVA) valve annuli, are chosen to drive the registration. This mapping technique is suitable for cardiac interventions, as it does not significantly lengthen procedure time, the selected valvular structures are easily identifiable in both datasets, and they also ensure a good anatomical alignment in the surrounding regions, enabling us to employ this technique for a variety of image-guided intracardiac interventions.

[1] VMTK: http://www.vmtk.org

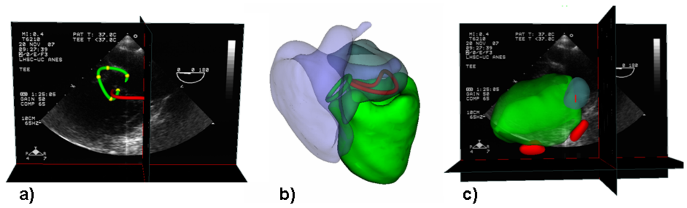


Fig. 2. a) Intra-operative US image, and b) pre-operative model, showing the MVA and AVA; c) Pre- and intra-operative datasets fused via feature-based registration. *Figure adapted from Linte et al. SPIE Medical Imaging 2008.*

Surgical Tool Tracking The surgeon must to know the position and orientation of the instruments with respect to the intrinsic surrounding anatomy at all times during the procedure. We integrate this feature via surgical tool tracking using the NDI Aurora (Northern Digital Inc., Waterloo, Canada) magnetic tracking system. As an example, for a MV implantation, we need virtual representations for the TEE transducer, valve-guiding tool and valve-fastening tool (see **Fig. 3**). 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.

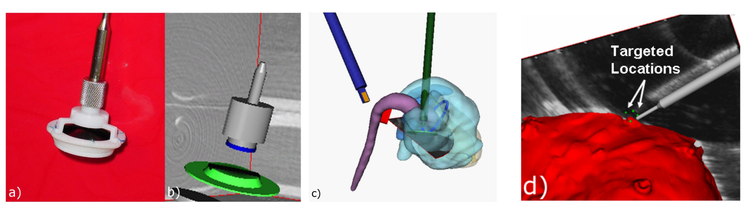


Fig. 3. a) Physical and b) virtual representation of a prosthetic mitral valve attached to guiding tool; c) *In vivo* AR environment; d) *in vitro* beating heart phantom study. *Figure adapted from Linte et al. Computer Aided Surgery 2008.*

Assessing Therapeutic Feasibility

Endocardial Interventions in Phantoms Our *in vitro* experiments mimicked left or right atrial endocardial procedures, where the surgeons would use a tracked instrument (i.e., ablation catheter) to "deliver therapy" to intracardiac targets in absence of direct vision. Moreover, we compared AR-guided targeting accuracy and procedure duration with those achieved under US image guidance alone (i.e., typical guidance modality employed for similar procedures), and endoscopic guidance (i.e., positive control, similar to direct vision) (**Fig. 4**).

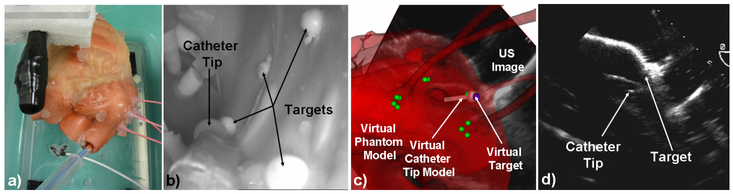


Fig. 4. a) Experimental setup; Endoscopic view of endocardial surgical targets; Model-enhanced US guidance view; Typical US image used for guidance. *Figure adapted from Linte et al. MICCAI 2008.*

Three users conducted the *in vitro* catheter navigation on four surgical targets, whose positions were tracked simultaneously using 5 DOF magnetic tracking sensor coils. Each user attempted the targets in 4 trials (i.e., 4 consecutive target sequences, each with a different target order) under each guidance modality (endoscopic, VR-US, and 2D US image guidance). The procedure outcome was assessed according to the targeting accuracy (**Fig. 5**).

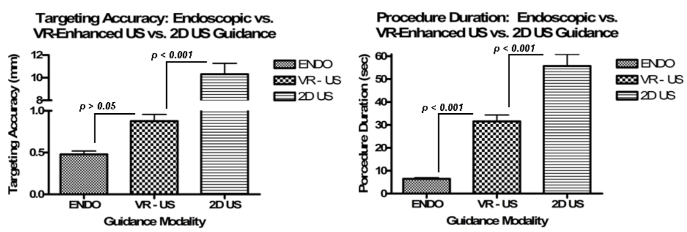


Fig. 5. Targeting accuracy and procedure duration achieved under endoscopic, VR-US, and 2D US image guidance, respectively. Note a significant improvement in both targeting accuracy and procedure duration under VR-US guidance with respect to 2D US image guidance. *Figure adapted from Linte et al. MICCAI 2009.*

Initiating Clinical Translation Direct access to the cardiac chamber was achieved using the Universal Cardiac Introducer (UCI) [5]. 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, in this case a laparoscopic clip applier. The procedure involves the positioning and fastening of the valve onto the native mitral annulus. Both steps entail navigating 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. 6**). Intra-operative assessment using Doppler imaging confirmed a successful valve placement, also observed in the post-procedure analysis.

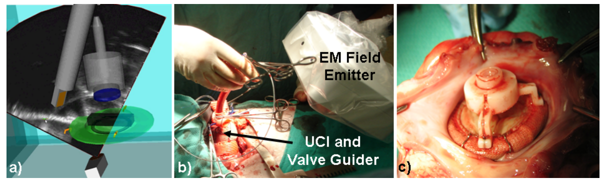


Fig. 6. a) AR environment showing virtual models of the US probe and surgical tools; b) OR setup during AR-guided interventions; c) Post-procedure assessment image. *Figure adapted from Linte et al. CARS 2009.*

2.2 Anesthesia: Spinal Facet Joint Injections

After gaining familiarity with our AR environment for intra-cardiac surgeries, our collaborating anesthetist suggested applying this augmented environment to a variety of problems faced regularly by anesthetists, such as facet joint injections and peripheral nerve blocks.

Lumbar facet joint injections are a good example of a particularly challenging therapy given the small, narrow channel between vertebrae, the oblique entry angle, relatively deep location and proximity to nerve tissue. The lumbar facet joint is a source of chronic pain for between 15% to 45% of patients with chronic lower back pain [7]. Most facet injections are performed using fluoroscopy or CT guidance, which is relatively expensive compared to US, and involves delivering radiation dose to the patient and health care providers.

Nerve blocks are ubiquitous in pain clinics as a means to avoid general anesthesia, acute pain control and nerve pain caused by cancer. Needle guidance for peripheral nerve blocks is limited to real-time 2D ultrasound, with no pre-procedural imaging, though nerve stimulation can sometimes be used to verify needle tip location. In most applications, the practitioner identifies the target tissue in the US image and then inserts the delivery needle in the US beam plane to approach the target. Given the non-uniform nature of US beam thickness, it is possible for the needle to appear in-plane but in fact be on an oblique angle with the needle tip several millimeters out of plane. This can be a particularly problematic issue, particularly for inexperienced residents.

Both of these applications, peripheral nerve blocks and facet injections, lend themselves well to our general system of augmenting US with virtual models of tracked tools. Our first goal was to assess two image guidance delivery systems, one making use of anatomical models, the second making use of only tracked tools (needle and US transducer). We chose the facet injection therapy as our first application since it is the more difficult procedure of the two. If the integration of VR tools with 2D US was adequate for image guidance to the facet joint then a "one size fits all" approach could be taken to both applications since peripheral blocks are already done exclusively with US guidance.

Evaluation of AR Guidance Platforms We performed a spine phantom study with the assistance of four anesthetists and four anesthesia residents. The clinicians evaluated three different guidance systems (**Fig.7**) to place a tracked needle on target for facet therapy injections in a lumbar spine phantom:

- US only (GS_{US}) — users were restricted to visualization provided via the display of the US scanner. They were free to manipulate the location and angle for the US transducer;
- US plus virtual tool representations (GS_{tools+}) — the users had full use of ultrasound, plus representations of the US and needle tool: the AR display consisted of the normal fixed view of the US fan with 'top' and 'side' views showing the angle and tip of the needle relative to the US beam in a manner similar to a standard technical drawing;
- Our complete 3D AR system (GS_{model+}) — the users had full use of the US, representations of US transducer and needle, plus an anatomical model of the spine based on a high resolution CT. The user had full control of their view of the virtual scene. The user was also able to define two "standard views" to toggle between: one view along the needle axis for trajectory and the second orthogonal to the first, for depth assessment.

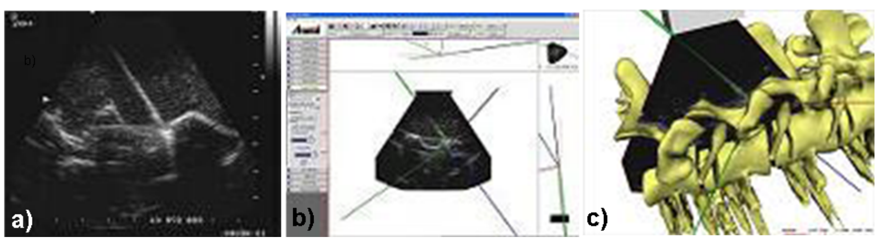


Fig. 7. The three guidance systems evaluated: a) US only (GS_{US}); b) US plus virtual tool representations (GS_{tools+}); c) Complete 3D AR system (GS_{model+}). The virtual needle tool also has x, y and z axes extensions with 10 mm markers to help locate the tip and its trajectory. *Figure adapted from Moore et al. MICCAI 2009.*

Details regarding the methodology and results of these experiments can be found in [8]. In summary, the RMS (root mean squared) distance error for GS_{US} was 10.22 mm, GS_{tools+} was 8.45 mm and GS_{model+} was 0.57 mm. **Fig. 8** shows a graphical representation of the results. These results strongly suggest a robust AR environment for 2D US will make US needle guidance a successful clinical practice for anesthesia needle deliveries. Further studies are planned in the area of peripheral nerve blocks to assess whether AR guidance can be of benefit.

Translation to the Clinical Environment A proof of concept cadaver study was performed using the GS_{model+} guidance system to assess the system in a clinical setting. Seven homologous points representing the spinal and lateral processes in the region of interest were defined in the pre-procedural CT and

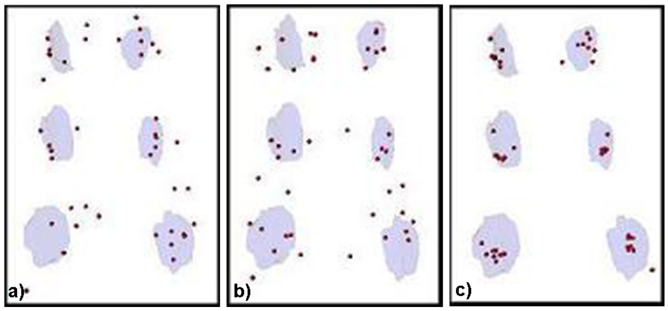


Fig. 8. Distribution of needle placements in coronal view: blue clouds are the facet joint targets, spheres are needle delivery locations. (a) US alone; (b) GS_{tools+} system, (c) GS_{model+} system. *Figure adapted from Moore et al. MICCAI 2009.*

interactively in the OR using 2D US to perform the point based registration. This procedure took approximately five minutes. To properly mimic a clinical therapy, radio opaque dye was injected when the user believed the needle was in the correct position. An anesthesia resident performed the needle delivery using the GS_{model+} guidance system, placing radio opaque dye at the left and right L2-3 and L3-4 facet joints. Accuracy was assessed independently by a radiologist using a post-procedure CT (**Fig.9**). The radiologist report stated signal enhancement (due to injected radio opaque dye) in and surrounding facet joints right and left, L2-3 and L3-4. This correlates with a clinically significant injection. There were no significant problems introducing the MTS into the clinical environment; care was taken to minimize the presence of ferro-magnetic materials in the immediate vicinity of the system, but no other special requirements were involved.

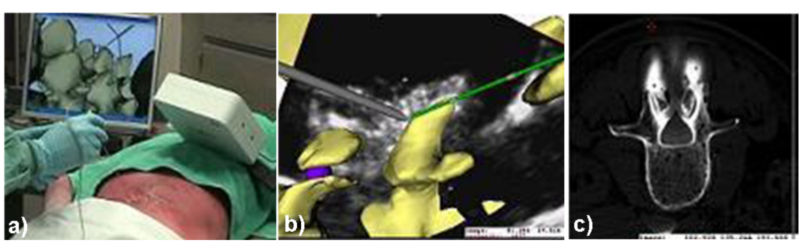


Fig. 9. Cadaver study and results. (a) OR environment showing AR scene, delivery needle and MTS field generator; (b) Close-up of GS_{model+} system including needle (gray), US fan and projected y axis (green); (c) post-procedure CT showing radio opaque dye at the L2-3 facet joints. *Figure adapted from Moore et al. MICCAI 2009.*

3 Discussion

We have summarized our experience in developing, evaluating and employing an AR-assisted navigation platform for use in both cardiac interventions, as well as facilitating anesthetic delivery. In addition to our *in vitro* experience, the translation into the clinic has helped us identify caveats specific to the OR environment that had not posed major concerns in the laboratory.

Accuracy is top concern in image-guided interventions! However, one must keep in mind that any IGS (image guided surgery) system will always be prone to inaccuracies arising from the difficulties in building perfect anatomical models and the limited accuracy of registering the models to the intra-operative patient anatomy. We quantified the alignment of the virtual models with the US images [9, 10], assessed the accuracy in targeting static and dynamic "surgical targets" in a beating heart phantom [11, 12], and evaluated the accuracy of "delivering anesthetic" to fact joints [8]. Given these applications, our system proves clinically feasible within the 1-2 mm accuracy limitations imposed by the MTS, and bearing in mind the scope of the pre-operative models to enhance intra-operative anatomical context. While greater overall accuracy is not highly required, better accuracy in the surgical target region is crucial.

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. For example, the cardiac project required us to build various prototypes for the valve- or ASD patch-guiding tools, embed the MTS sensors inside the tools, adapt a laparoscopic stapler for abdominal interventions as a fastening device for both the valve and ASD patch, and test their compatibility with the surgical environment. Most off-the-shelf tools do not comply with our requirements, and the most efficient approach would be to involve a medical device manufacturer into the project.

An OR is often too "busy" to allow the use of an optical tracking system given its line of sight restrictions, hence raising the need to employ magnetic tracking technology, which in turn implies the avoidance of any ferromagnetic objects in close proximity to the magnetic field emitter [13]. To avoid obstruction of both surgical work space and work flow, we embedded the magnetic field generator within the mattress of the operating table, underneath the "region of interest"; as such, tracked sensors closer to the field generator than large bodies of metal (i.e., rib spreader), making the MTS substantially more robust.

Minimal invasiveness restricts both visual and surgical access; for both applications described, physicians can't "see" what they do; the AR guidance platform **is** their eyes and we are well aware of the efforts and time required to make its way into conventional surgery. Over the years, surgeons have become familiar with "standard views" of human anatomy. Although we are now able to provide an unlimited range of views of tools and surgical targets, it is often best to stick to the displays surgeons are most familiar with. The best approach is to make the new look as much like the old: give them the time to get accustomed to the new environments, let them tell you how impressed they are, rather than overwhelming them with new technology at once!

4 Conclusions

Our results indicate that AR assisted ultrasound guidance for both intracardiac therapy as well as facet injections shows tremendous promise for increasing patient safety and comfort, greatly reducing or even eliminating radiation dose to both patient and clinicians, as well as reducing health care costs. Additional studies will be required to assess the value of AR techniques to augment real time US for other applications, such as peripheral nerve block therapies, neurosurgical applications, and abdominal interventions.

References

1. Caudell, T.P.: Introduction to augmented and virtual reality. Volume 2351 of Proc. of SPIE - Telemanipulator and Telepresence Technologies. (1994) 272–81
2. 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
3. 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
4. Wilson, K., Jones, D.L., Linte, C.A. *et al.*: Dynamic cardiac mapping on patient-specific models. In: Lect Notes Comput Sci. Volume 5241 of Proc. MICCAI. (2008) 967–74
5. 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
6. Marmurek, J., Wedlake, C., Pardasani, U., Eagleson, R., Peters, T.M.: Image-guided laser projection for port placement surgery. Volume 119 of Stud Health Technol Inform. (2007) 367–72
7. Boswell, M., Colson, J., Epter, R. *et al.*: A systematic review of therapeutic facet joint interventions in chronic spinal pain. Pain Physician **10** (2007) 229–253
8. Moore, J. *et al.*: Image guidance for spinal facet injections using tracked ultrasound. In: Lect Notes Comput Sci. (2009)
9. Linte, C.A., Moore, J., Wedlake, C.*et al.*: Inside the beating heart: An *in vivo* feasibility study on fusing pre- and intra-operative imaging for minimally invasive therapy. Int J CARS **4** (2009) 113–122
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., Moore, J., Wiles, A.D. *et al.*: Targeting accuracy under model-to-subject misalignments in model-guided cardiac surgery. In: Lect Notes Comput Sci. Volume TBA of Proc. MICCAI. (2009) In Press
12. Linte, C.A., Moore, J., Wiles, A.D.*et al.*: In vitro cardiac catheter navigation via augmented reality surgical guidance. In: Proc. of SPIE. - Medical Imaging. Volume 7261. (2009) 72610O–9
13. 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

Development of Endoscopic Robot System with Augmented Reality Functions for NOTES that Enables Activation of Four Robotic Forceps

Naoki Suzuki[1], Asaki Hattori[1], Kazuo Tanoue[2], Satoshi Ieiri[2], Kozo Konishi[3], Hajime Kenmotsu[2], Makoto Hashizume[2,3]

1 Institute for High Dimensional Medical Imaging, The Jikei University School of Medicine, Tokyo, Japan, 2 Dept of Advanced Medicine and Innovative Technology, Kyushu University Hospital, Fukuoka, Japan, 3 Dept of Future Medicine and Innovative Medical Information, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan
{nsuzuki, hat}@jikei.ac.jp, {tanoue_k@dem, satoshi@, konizou@surg2, kenken@dem, mhashi@dem}.med.kyushu-u.ac.jp

Abstract. The purpose of this research project is to develop an endoscopic surgery robot that can carry out a surgery in the abdominal cavity using the same surgical technique as that of an open abdominal surgery by mounting various wire-driven manipulators at the tip of the scope of an endoscope. We have been researching and developing: forceps-type manipulator that can open and close, and move up and down by a wire driven mechanism, its drive mechanism and its control software. We report on the progress of the development of the above, and the surgical robot that is loaded with an endoscope that has 4 manipulators at its tip. This is so that 2 surgeons can use 2 manipulators each simultaneously in the surgical process.

Keywords: Endoscopic surgery, Robotic surgery, NOTES

1. Purpose

We have been developing an endoscopic surgery robot for abdominal surgery since 2001. The robot has a lens at the tip of the endoscope that acts as an eye, and is equipped with two forceps-type manipulators that function as right and left hands. It is inserted via the mouth and reaches the stomach. This system can operate via the same surgical process as an open abdominal surgery in the stomach. We are currently developing a system that enables NOTES (Natural Orifice Transluminal Endoscopic Surgery) that penetrate the stomach and operate on the internal organs in the abdominal region.

We report on the development of a system that enables two surgeons each to use both hands so that they can jointly carry out a surgery in the abdominal cavity, by adding two more manipulators to the system, to possess a total of four manipulators. We also report on a wire tensile control-type manipulator control mechanism that effectively controls the forceps of the robot.

2. Method

2-1 Manipulator Drive Mechanism

The forceps-type manipulator for this system uses a wire-driven mechanism so that it does not destroy the flexibility of the endoscope. In this way, the system moves the manipulator and opens and closes the forceps. The manipulator is shown in figure 1. It has a mechanism for the forceps at its tip and can move up and down, right and left under the control of three wires. It can also open and close the forceps under the control of one wire. We made two pairs of manipulators were mounted at the tip of an endoscope, as shown in figure 2. The four manipulators can be used by two surgeons, who each control right and left manipulators. The coordinated actions of these manipulators enable complicated surgical operations to be performed.

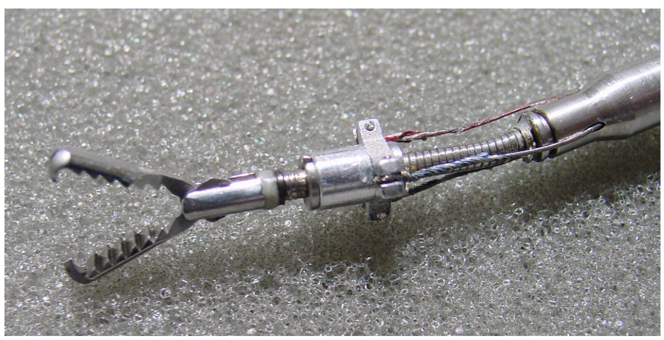


Fig. 1. Tip of Manipulator

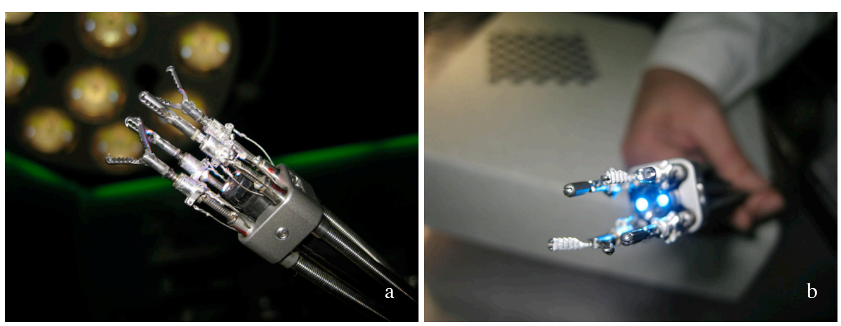


Fig. 2. Tip of Endoscope Mounted with 4 Manipulators

Figure 3 shows the wire-driven mechanism of the manipulator. The drive mechanism consists largely of two parts. One drive mechanism drives two manipulators.

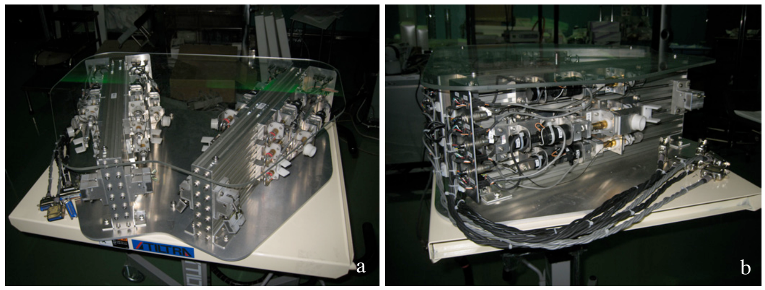


Fig. 3. Tractile Force Control Type Wire Drive Mechanism

On the right and the left sides of one drive mechanism, we positioned a stepping motor to wire-drive manipulators and four power sensors that measure the traction force of each wire.

Figure 4 shows the console for the surgeons to control the manipulators. The console enables surgeons to activate two manipulators up and down, right and left, and also enables the forceps to be opened and closed. We have created one more console so that four manipulators can be used simultaneously.

Fig. 4. Using the Console

2-2 Manipulator Drive Mechanism

We have designed software to control the traction force of the wire so that the above-mentioned drive mechanism can effectively drive the manipulators.

We designed software to control and lessen as much as possible the change in form of the endoscope when the robot is inserted into the abdominal cavity, and the change in the traction force of the wire when the manipulator is opened and closed. This was done by monitoring the traction force of the wire that drives the manipulator by the power sensor positioned on the wire drive mechanism. We also added two other functions: one to release the wire when an abnormally large force is applied to the wire when the system is activated, and one to alert the surgeon to any disconnection of the wire. We also designed a function that would alert the surgeon when the tractive motor driving the wire stops rolling due to any abnormal circumstances.

In addition, we created a function so that reveals to the surgeon the hardness of any tissues grabbed by the manipulator, by measuring the traction force of the wire used to open and close the forceps and detecting the change in traction force when grabbing tissues. Figure 5 shows the measuring of the traction force and the resultant display of hardness.

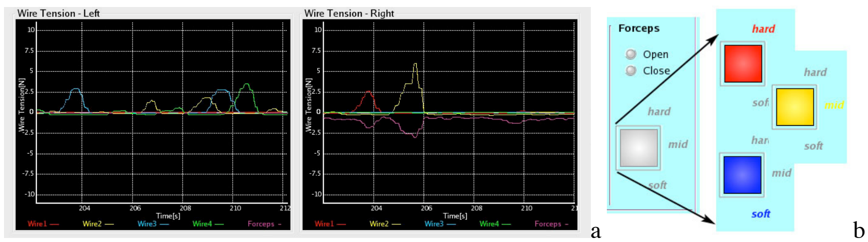


Fig. 5. Measurement of Tensile Force (a) and the Display of Results of Hardness (b)

2-3 Augmented Reality Function

We developed a prototype of real-time information integration and display system for endoscoipc robot surgery by detecting the 3D position and direction of the tip of the endoscopic robot to superimpose the patient's ultrasound image onto the endoscopic image. (Figure 6) We have made it to function so that the position that is marked on the ultrasound image (the red cross in Figure 6a, b) is also displayed on the superimposed image screen.

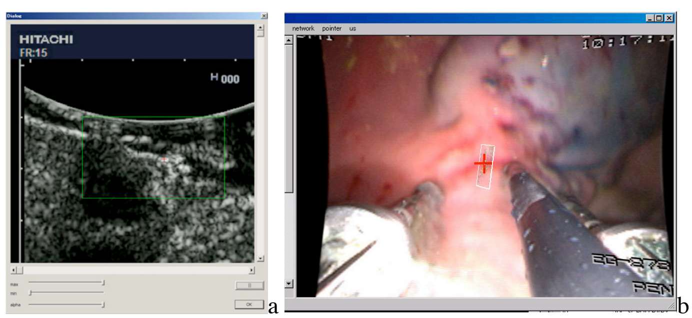


Fig. 6. Display screen of the system. a: Selecting the region of interest (green) on the captured ultrasound image and marking the part of attention (red) b: Superimposing the selected region of the ultrasound image onto the endoscopic image and displaying the marked part of attention

3. Results and Conclusions

We performed animal experiments to test the system, as shown in Figure 7. The figure shows two surgeons operating on a anesthetized pig (weight: 40 kg.). The surgeons have inserted the endoscopic robot into the pig. A console and a monitor displaying endoscopic images were positioned in front of each surgeon. In this experiment, we examined, among other things, whether the surgeons could handle gastric mucosa properly, and whether they could carry out dissection and excision of the gastric mucosa.

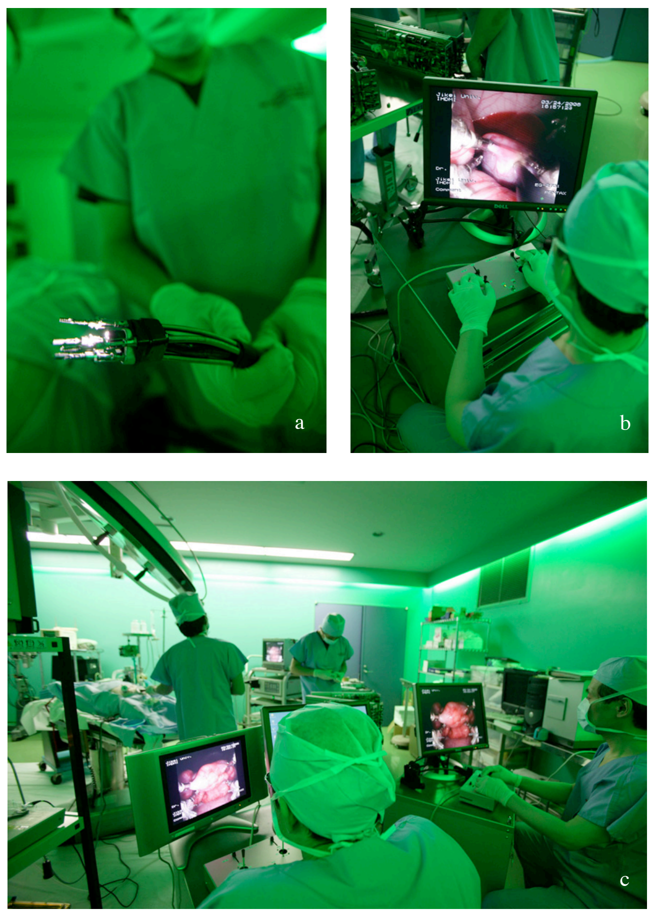


Fig. 7. Animal Test to Experiment the System

We also considered whether an X-ray CT device could effectively monitor the state of the robot inserted into the gastrointestinal tract of the animal during the experiment, and monitor the inner structure of the manipulator, including the activation wire, simultaneously with the state of the gastrointestinal tract wall. Figure 8 shows volume rendering images of the tip of the endoscopic robot loaded with four manipulators, as measured by an X-ray CT device.

By choosing the measurement conditions for X-ray CT and using appropriate 3D display methods, we were able to confirm not only the positions of the four manipulators and the states (open or closed) of the forceps, but also the states of the wires that drive the manipulators.

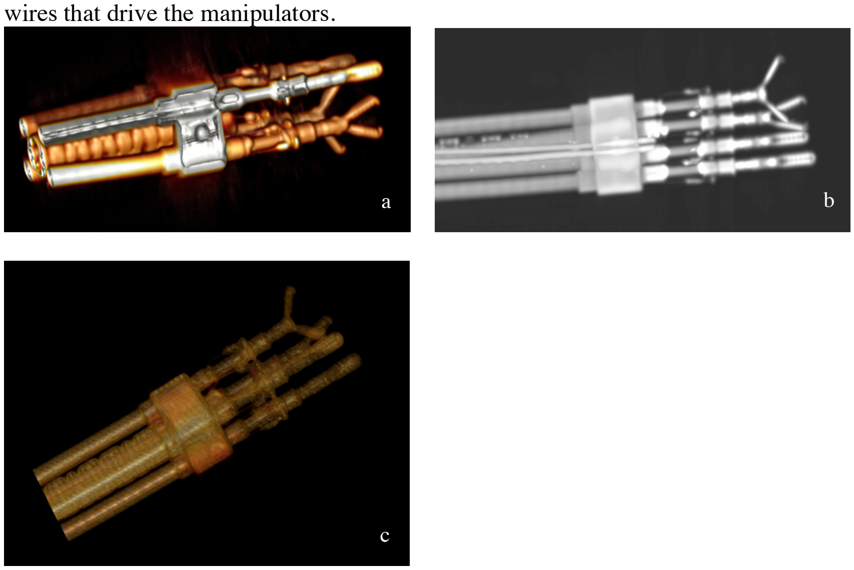


Fig. 8. Result of Measurement by X-ray CT Device of the State of Tip of Endoscopic Robot and the Inner Structure Including Drive Wire

Based on these results, we plan to develop an advanced endoscopic robot that can carry out NOTES surgical procedures under an environment close to that of open abdominal surgery by equipping the tip of an endoscope with one to five surgical instruments of various kinds, according to the surgical situation.

Acknowledgment

This study was funded by NEDO Intelligent Surgical Instruments Project.

References

1. Salisbury J.K.: The heart of microsurgery. Mechanical Engineering Magazine, ASME Int'l., 120, 47--51 (1998)
2. Reichenspurner H., Damiano R.J., Mack M., Boehm D.H., Gulbins H., Detter C., Meiser B., Ellgass R., Reichart B.: Use of the voice-controlled and computer-assisted surgical system

ZEUS for endoscopic coronary artery bypass grafting. J Thorac Cardiovasc Surg, 118, 11-6 (1999)
3. Guthart G.S., Salisbury J.K.L: The Intuitive Telesurgery System: Overview and Application. Proc. of the IEEE International Conference on Robotics and Automation (ICRA2000). San Francisco, (2000).
4. Piskun, G., Rajpal, S.: Transumbilical laparoscopic cholecystectomy utilizes no incisions outside the umbilicus. J Laparoendosc Adv Surg Tech A. 9, 361--364 (1999)
5. Kalloo, A.N., Singh V.K., Jagannath S.B., Niiyama H., Hill S.L., Vaughn C.A., Magee C.A., Kantsevoy S.V.: Flexible transgastric peritoneoscopy: a novel approach to diagnostic and therapeutic interventions in the peritoneal cavity. Gastrointest Endosc. 61, 601--606 (2004)
6. Kantsevoy S.V., Hu B., Jagannath S.B., Vaughn C.A., Beitler D.M., Chung S.C.C., Cotton P.B., Gostout C.J., Hawes R.H., Pasricha P.J., Magee C.A., Pipitone L.J., Talamini M.A., Kalloo A.N.: Per-oral transgastric endoscopic splenectomy: is it possible?. Surg Endosc. 20, 522--525 (2006)
7. Marescaux J., Dallemagne B., Perretta S., Wattiez A,. Mutter D., Coumaros D.: Surgery Without Scars: Report of Transluminal Cholecystectomy in a Human Being. Arch Surg. 142, 823--826 (2007)

A Practical Approach for Intraoperative Contextual In-Situ Visualization

Christoph Bichlmeier[1], Maksym Kipot[1], Stuart Holdstock[1], Sandro Michael Heining[2], Ekkehard Euler[2], and Nassir Navab[1]

[1] Department of Chair for Computer Aided Medical Procedures & Augmented Reality, TU Munich, Germany
[2] Trauma Surgery Department, Klinikum Innenstadt, LMU, Munich, Germany

Abstract. The composition of real and virtual entities has a major contribution to the usability and acceptance of an Augmented Reality (AR) Scene. The composition method highly depends on the accessibility of information in order to allow the user to intuitively perceive relative and absolute distances of objects within the AR scene. This paper proposes a novel, flexible and practical approach of creating a context preserving view onto the operation site. It exclusively employs information from video data of video see-through AR systems to compute context information.

1 Introduction

Providing a reasonable composition of real and virtual entities within AR environments is a crucial task and a hot subject of recent published research [1–7]. In particular, the composition becomes important when virtual objects are physically positioned behind real ones. This kind of object topology is given when medical imaging data such as CT or MRI is visualized in-situ, however, appears to be floating above the patient's body [2, 8–12]. The previously introduced *Contextual In-Situ Visualization* [3, 4] for AR was inspired by *Focus & Context Visualization* methods having been previously introduced for virtual worlds [13–15]. These approaches use primarily the geometry of the patient's skin generating context information for better perception of deeper seated anatomy. However, they are not practical and flexible enough for being intraoperatively applied in future AR systems. The majority of image guided surgeries do not use imaging data that is large enough to cover the skin geometry. Even if the skin geometry is available, we deal here with only static context information captured at a certain point in time. For this reason, the skin serving as the major context layer cannot adapt to geometry deformation or color changes.

We decided to follow the approaches of Kalkofen et al. [1], Stoyanov et al. [2] and Avery et al. [6] using exclusively information from video data of video see-through devices to generate context information. We extended these approaches with additional visual cues reacting on changing light and color conditions, deformation, and interaction with surgical instruments. This paper presents the rendering pipeline of a new version of *Contextual In-Situ Visualization* strongly optimized for surgical AR environments. In addition, we discuss our first approach of designing of an experimental setup to qualitatively assess visualization methods that are intended for improving the spatial perception of medical AR environments.

2 Method

After the description of the used AR system, we introduce a new method for contextual in-situ visualization and our preliminary experimental setup up for evaluating the quality of depth perception.

2.1 In-Situ Visualization

The rendering approach has been tested with the RAMP system introduced by Sauer et al. [8] as well as with a new video see-through head mounted display (HMD), which is an NVIS NVisor SX having a 1280x1064, 60Hz display with 24-bit colors. A black and white 640x480 resolution PTGrey Flea camera is used for the infrared tracking. Two PTGrey Flea color cameras with a resolution of 1024x780 capture the view in front of the HMD and these images are displayed in the HMD simulating the view from a user's eyes. The HMD is tracked using both an inside-out tracking as well as an outside-in tracking system. The inside-out tracking system uses the infrared camera attached to the HMD to track a set of infrared markers fixed to an arc and this allows head pose estimation to be calculated. This tracking setup provides high rotational accuracy. The outside-in tracking system is from the company A.R.T. GmbH, Germany and we have four infrared cameras fixed on a ceiling mount with each camera positioned in a corner to enable a clear view of the infrared markers from all angles. Infrared markers attached to the HMD enable an outside-in tracking which is used as a backup system for head pose estimation and only in the case when the inside-out tracking system fails. However, an advantage of this setup is that the HMD user can move freely within the AR scene even when the infrared markers on the arc are out of the field of view of the HMD's infrared camera and it also provides for additional image stability. This dramatically increases the usability of the system even though the outside-in tracking cannot provide such high rotational precision [16]. In our preliminary but extendable AR setup an endoscopic instrument and an object simulating the patient's body are tracked.

2.2 Rendering Pipeline

Our visualization pipeline installs a transparent vision channel [17, 4] virtually onto the deepest context layer of the AR scene, which is in our case the skin, that is aligned with the line of sight of the HMD user. A transparency fading effect is applied at the border of the vision channel in order to generate a smooth transition from the original skin color of the video to the assigned transparency of the vision channel.

$$AF = ((1.0 - AL)/FR) * (CPVC - VCR + FR) + AL; \tag{1}$$

AF describes the alpha value within the region of transparency transition. It is calculated with the parameters of the initial global alpha value inside the vision channel

(AL), the width of the concentric disk describing the region of transition (FR), the current position of the pixel within the vision channel being used for the caluclation of the alpha value(CPVC), and the vision channel radius (VCR). The proposed rendering pipeline for our method of image composition involves four main components using information from tracking data and video images. The transparency generating the context information within the vision channel is parameterized with the following functions on the video data:

- MATERIAL1: A direction invariant Sobel filter for the detection of edges in video images caused by surgical instruments and trocars, surgeons' hands, the open operation site itself, blood drops, birthmarks, scars, cover sheeting, artificially introduced marks by a sterile pen onto the skin.
- MATERIAL2: Detection of specular highlights caused by bright metallic light reflecting instruments, humid skin surface, fluids, bright/specific colors in general.

In addition, two combined shadow effects enhance the spatial appearance and provide interactive visual feedback:

- SHADOW1: The combination of MATERIAL1 and MATERIAL2 is projected shadow-like onto virtual objects behind the skin layer using OpenGL supported projective texture mapping.
- SHADOW2: Shadow mapping [18] applied to all virtual objects.

Video data coming from a video capturing device like an HMD, a laparoscope or an arthroscope camera is processed by a GLSL [3] based shader. For modular computation and final composition, Frame Buffer Objects (FBO) for GPU accelerated off-screen rendering have been employed. Regarding MATERIAL1 (see Fig. 1(a)), the transparency (AE) value of each pixel of the transparency map is computed by comparing the gray values (GV) of both the original image pixel (OP) and the extracted edge pixel (EP) and adjusting it with a transparency level (AL):

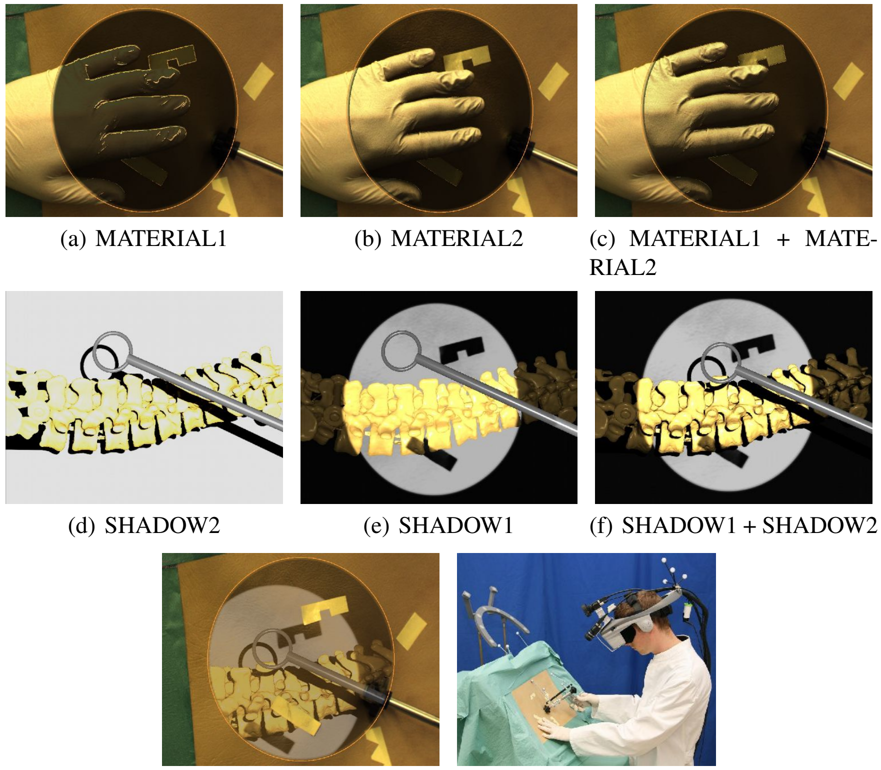
$$AE = OP.GV/EP.GV + AL \tag{2}$$

MATERIAL2 thresholds the gray scale version of the video data to detect brightly colored regions (GV), which remain opaque while darker regions are coded with a high transparency value:

$$AS = AL + 1.0 - (1.0 - OP.GV)/(1.0 - ST) \tag{3}$$

The alpha value (AS) of a pixel of the video image uses the following parameters. OP is the original pixel value, AL is again the initial global alpha value inside the vision channel and ST determines the threshold for bright regions (see Fig. 1(b)). MATERIAL1 and MATERIAL2 are then combined as shown in Fig. 1(c). The result is then forwarded to another shader program performing projective texture mapping [18] to feed SHADOW1. SHADOW1 and SHADOW2 apply *percentage closer filtering* (PCF) [19], which is supported by the used graphics hardware [4] to create smooth shadow and avoid

[3] OpenGL Shading Language
[4] Nvidia™Geforce 8800 Ultra



(a) MATERIAL1 (b) MATERIAL2 (c) MATERIAL1 + MATERIAL2

(d) SHADOW2 (e) SHADOW1 (f) SHADOW1 + SHADOW2

(g) Frame Buffer (h) One of 26 subjects.

Fig. 1. Full pipeline described with screenshots of an exemplary AR scene.

aliasing artifacts (see Fig. 1(d)). The method of combining SHADOW1 and SHADOW2 follows a natural shadowing principle based on *diffusion* and *interference*. The blurriness of a shadow increases and the contrast decreases in proportion to an increase of the distance between a light source and a fixed object in space. According to our assumption that 3D objects are located behind the context layer, features filtered by MATERIAL1 and MATERIAL2 can be projected with SHADOW1 onto these objects so that their shadows will be sharper and have more contrast. Blurriness can be adjusted with PCF kernels and both, SHADOW1 and SHADOW2, can be weighted. However, blurring and weighting of SHADOW1 is based on an approximation of the distance between the skin surface and virtual objects behind the skin since the 3D pose of the skin is not known. The light for calculating SHADOW1 and SHADOW2 is positioned relative to the HMD of the user in our case $20cm$ to the right.

At the end of the rendering pipeline we deal with two textures called *Processed Video* and *Final Virtual Scene*. *Processed Video* represents the video data with a previously individually computed alpha value for each pixel within the vision channel. The other texture contains all virtual 3D information including the composition of SHADOW1 and SHADOW2. *Processed Video* and *Final Virtual Scene* are then composed with OpenGL blending and rendered into the Frame Buffer resulting in the final AR scene (see Fig. 1(g)).

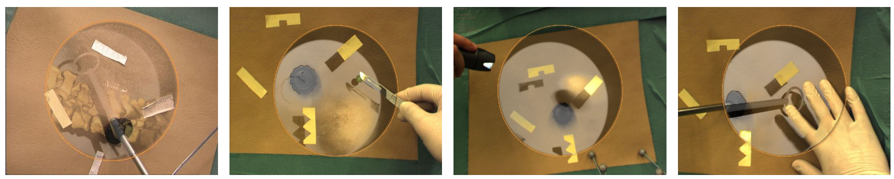
2.3 Assessing the Perceptive Quality

User studies are essential to qualitatively and quantitatively prove the advantage of a proposed method. However, designing an experiment that satisfies the quality criteria objectivity, reliability and validity is a difficult task. For assessing the perceptive quality to estimate the relative or absolute depth of objects in the designated medical AR scene, in particular the decision on the right measuring instruments is difficult. The hardware weight, lens configuration of the cameras, FOV of the cameras and display, tracking quality, possible stability, resolution, color depth and update rate, presence or absence of medical knowledge, vision abilities of subjects and many more parameters may bias the measuring data.
Here, we present the status quo of an iteratively tested and improved evaluation environment that has been designed to assess the quality of different methods for image composition for instance our proposed *Contextual In-Situ Visualization* (see Fig. 2). The experiment requires subjects to insert a tracked endoscopic instrument through a trocar to reach target positions behind a synthetic skin layer (see Fig. 1(h)). The instrument is superimposed with its virtual counterpart that has been augmented with a virtual ring at the instrument tip. The subject's task is to guide the ring of a set of virtual, unicolored, randomly sized and positioned balls that are physically located behind the skin layer. The ball changes its color in case of collision. In addition, to the number of collisions, we measure time of performance, time of collision, number of tries (according to the protocol the ring has to be guided straight forward over the ball), path length of the instrument tip, time it took to first reach the area near the ball as measured from a start position, and path length of the instrument tip along the instrument axes (motion into depth).

3 Results

Figure 2(a) shows an endoscopic instrument penetrating a previously inserted trocar above a simulated operation site for key hole surgery. The skin around the trocar shows in this case only homogeneous color distribution providing almost no information for MATERIAL1 and MATERIAL2. However, guidance of instruments causes deformation of the skin and influences the color brightness of the skin. This reactively changing light reflection can be detected by MATERIAL2 to generate reactive transparency changes. In order to feed MATERIAL1, we attached small glue strips having a slightly different color than the skin (see Fig. 2(a)). For intra operative surgical planning in some cases sterile markers are used to draw rough sketches of the hidden anatomy onto the patient's skin. MATERIAL1 and MATERIAL2 can be configured to detect also the color of such sterile markers to enable additional context information. Alternatively light patterns can be projected onto the skin. Figure 2(c) shows a hand held flashlight causing a bright spot on the skin that is detected by MATERIAL2. This sterile light spot again causes partial opacity of the vision channel. In addition, its shadow is thrown onto the virtual anatomy and a virtual plane in the background of the scene. SHADOW1 can produce shadow from any object that is shown in the video image. This includes the surgeon's hand (see Fig. 2(d)) and surgical instruments such as a scalpel (see Fig. 2(b)) or the endoscopic instrument with its attached marker tree for optical tracking. The effect of masking the scene with the user's hand has been previously shown with distinguishable blueish gloves having an exclusive color in the AR scene [3]. The current configuration of MATERIAL1 can also deal with skin colored gloves that are mainly used in todays ORs (see Fig. 1(c)).

Although, we got positive feedback from the majority of 26 subjects that tested the



(a) Deformation of the skin (b) Extracted instruments cast shadow (c) Flashlight causes context information. (d) Hand causes shadow feedback.

Fig. 2. Different features of the proposed image composition.

proposed method with our experimental setup described in section 2.3 our measuring data does not show significant differences when comparing our Contextual In-Situ Visualization with simple superimposed opaque or transparent balls onto an opaque skin. The positive feedback is derived from a questionnaire that had to be completed after the experiment. To the question *How supportive do you consider the shown visualization modes for your task*, the number answering either supportive or very supportive is 6 for the opaque balls, 4 for the transparent balls and 18 for our new method. With the question *How realistic do you find each visualization mode*, the number answering realistic or very realistic is 4 for the opaque balls mode, 4 for the transparent balls and 20 for our new method.

4 Discussion

Fig. 2(c) shows the generation of depth cues using a light source changing the illumination of the context layer, i.e. the skin. Instead of light being visible for the humans eye, infrared light can be used to illuminate the skin for MATERIAL2. Infrared light can be reflected by the skin and is visible for cameras. In this case the homogeneous illumination conditions in ORs are not disturbed. Currently parametrization of MATERIAL1, MATERIAL2, SHADOW1 and SHADOW2 is configured manually by the user. However, future work will address the dynamic parametrization of the proposed functions. This can involve the analysis of the color histogram of the video over time to decide on the thresholding and weighting of MATERIAL1 and MATERIAL2. Also the blurriness and contrast of shadow effects can be adjusted according to analysis of the current lighting conditions. A major part of future work will be the interdisciplinary improvement of the evaluation environment together with psychologist having a strong background in methodology to guarantee reliability, validity and objectivity of the study. The positive feedback from subjects strongly motivates us to study intensively whether the visualization method or the evaluation design causes the lack of significant measuring data.

5 Conclusion

This paper proposes a novel, flexible and practical approach of creating a context preserving view inside the patient during an operation. It exclusively employs information from video data from video see-through AR systems to compute context information, which allows for the processing of deformable context layers. The generated graphical effects from MATERIAL1 and MATERIAL2 that are part of the proposed visualization pipeline provide the depth cues of occlusion and motion parallax even though the illumination, color and geometry conditions completely change during the surgical procedure, e.g. the skin is cut or deformed, minimally invasive trocars are installed or an open operation is being performed. By adjusting parameters the masking of the virtual part of the scene with conventional surgical hand gloves can be achieved. The effects SHADOW1 and SHADOW2 can improve the 3D layout of the AR scene. In addition, they provide visual feedback when the user starts to interact, e.g. guidance of a surgical instrument, with the scene. The proposed method can be applied for any video see-through technology being used for AR applications. In the medical field this includes for instance laparoscope or arthroscope cameras. We strongly believe that also video see through HMDs used for medical procedures would benefit from our approach once they are integrated into the operating theater.

6 Acknowledgements

We would like to thank Frank Sauer, Ali Khamene, and Sebastian Vogt from Siemens Corporate Research (SCR) for the design, setup, and implementation of the RAMP system they provided us. Thanks also to Tobias Sielhorst having constructed the new HMD. We would also like to thank A.R.T. GmbH for providing cameras and software for the outside-in tracking system. We want to express our gratitude to the radiologists and surgeons of Klinikum Innenstadt in Munich for their precious contribution in obtaining medical data and evaluating our systems. Many thanks also to the crew at NARVIS lab.

References

1. Kalkofen, D., Mendez, E., Schmalstieg, D.: Comprehensible visualization for augmented reality. IEEE Transactions on Visualization and Computer Graphics **15**(2) (2009) 193–204
2. Stoyanov, D., Mylonas, G.P., Lerotic, M., Chung, A.J., Yang, G.Z.: Intra-operative visualizations: Perceptual fidelity and human factors. J. Display Technol. **4**(4) (2008) 491–501
3. Kutter, O., Aichert, A., Bichlmeier, C., Heining, S.M., Ockert, B., Euler, E., Navab, N.: Real-time Volume Rendering for High Quality Visualization in Augmented Reality. In: International Workshop on Augmented environments for Medical Imaging including Augmented Reality in Computer-aided Surgery (AMI-ARCS 2008), New York, USA, MICCAI Society (Sept. 2008)

4. Bichlmeier, C., Wimmer, F., Heining, S., Navab, N.: 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). (November 2007) 129–138
5. Swan, J.E.I., Livingston, M.A., Smallman, H.S., Brown, D., Baillot, Y., Gabbard, J.L., Hix, D.: A perceptual matching technique for depth judgments in optical, see-through augmented reality. In: VR '06: Proceedings of the IEEE Conference on Virtual Reality, Washington, DC, USA, IEEE Computer Society (2006) 19–26
6. Avery, B., S.C., Thomas, B.: Improving spatial perception for augmented reality x-ray vision. In: VR '09: Proceedings of the IEEE Conference on Virtual Reality, Louisiana, USA, IEEE Computer Society (2009) 79–82
7. Nicolaou, M., James, A., Lo, B., Darzi, A., Guang-Zhong, Y.: Invisible shadow for navigation and planning in minimal invasive surgery. (2005) 25–32
8. Sauer, F., Khamene, A., Bascle, B., Vogt, S., Rubinob, G.J.: Augmented reality visualization in imri operating room: System description and pre-clinical testing. In: Proceedings of SPIE, Medical Imaging. Volume 4681. (2002) 446–454
9. Feuerstein, M., Mussack, T., Heining, S.M., Navab, N.: Intraoperative laparoscope augmentation for port placement and resection planning in minimally invasive liver resection. IEEE Trans. Med. Imag. **27**(3) (March 2008) 355–369
10. Bajura, M., Fuchs, H., Ohbuchi, R.: Merging virtual objects with the real world: seeing ultrasound imagery within the patient. In: Proceedings of the 19th annual conference on Computer graphics and interactive techniques, ACM Press (1992) 203–210
11. Fuchs, H., A., S., Pisano, E., Garrett, W., Hirota, G., Livingston, M., Whitton, M., Pizer, S.: Towards performing ultrasound-guided needle biopsies from within a head-mounted display. (1996) 591–600
12. Wanschitz, F., Birkfellner, W., Figl, M., Patruta, S., Wagner, A., Watzinger, F., Yerit, K., Schicho, K., Hanel, R., Kainberger, F., Imhof, H., Bergmann, H., Ewers, R.: Computer-enhanced stereoscopic vision in a head-mounted display for oral implant surgery. Clinical Oral Implants Research **13**(6) (2002)
13. Bruckner, S., Grimm, S., Kanitsar, A.: Illustrative context-preserving exploration of volume data. IEEE Transactions on Visualization and Computer Graphics **12**(6) (2006) 1559–1569 Member-M. Eduard Groller.
14. Burns, M., Haidacher, M., Wein, W., Viola, I., Groeller, E.: Feature emphasis and contextual cutaways for multimodal medical visualization. In: EuroVis 2007 Proceedings. (May 2007)
15. Krueger, J., Schneider, J., Westermann, R.: Clearview: An interactive context preserving hotspot visualization technique. IEEE Transactions on Visualization and Computer Graphics **12**(5) (2006) 941–948
16. Hoff, W., Vincent, T.: Analysis of head pose accuracy in augmented reality. IEEE Transactions on Visualization and Computer Graphics **6**(4) (2000) 319–334
17. Pinna, B., Tanca, M.: Perceptual organization reconsidered in the light of the watercolor illusion: The problem of perception of holes and the object-hole effect. J. Vis. **8**(7) (5 2008) 1–15
18. Williams, L.: Casting curved shadows on curved surfaces. In: In Computer Graphics (SIGGRAPH '78 Proceedings. (1978) 270–274
19. Reeves, W.T., Salesin, D.H., Cook, R.L.: Rendering antialiased shadows with depth maps. SIGGRAPH Comput. Graph. **21**(4) (1987) 283–291

Session 2 – Efficiency in AR and Registration

Interactive 3D auto-stereoscopic image guided surgical navigation system with GPU accelerated high-speed processing

Huy Hoang Tran[1], Hiromasa Yamashita[1], Ken Masamune[1], Takeyoshi Dohi[1], and Hongen Liao[2]

[1] Graduate School of Information Science and Technology, The University of Tokyo, Japan,
[2] Graduate School of Engineering, The University of Tokyo, Japan

Abstract. We present an interactive image guided navigation system using the auto-stereoscopic Integral Videography (IV) with high-speed GPU processing. IV images produce full-colored, full parallax visualization of 3D data (CT data, surface model, etc.) in real space. These features eliminate the difficulties in defining the position and orientation of the interested objects or planning surgical paths, especially in orthopedic surgeries. Instead of being displayed on a separated screen, 3D images and related information are projected onto the patient using a half-silvered mirror attached to a 3D display to support surgeons with fast and intuitive navigation. We developed a hardware rendering method that implements the current IV rendering algorithm onto the Graphical processing unit (GPU). This implementation allowed IV images to be rendered at more than 10 frames per second in most cases. With our proposed system, surgeons can smoothly perform interactions like rotating and scaling with the IV image in almost real-time. This feature is particularly useful in surgical navigation systems where real-time vision of the surgical site is consistently required.

1 Introduction

Image guided surgery systems based on pre-operative image (CT, MRI) as well as intra-operative image (ultra sound, fluoroscopy) have been widely developed and have proven efficient in clinical cases [1, 2], especially in minimally invasive surgery (MIS). In most of current systems, navigation information is often displayed on a separated external 2D display, which does not give the surgeons direct and intuitive vision of the surgical area. This limitation may lead to difficulties in defining objects of interest or planning surgical paths, especially in orthopedic surgery [3].
Efforts in solving this problem include integrating the image overlay system, where 3D images are projected directly onto the real objects [4]. Another idea is introducing a stereo video-see-through head mounted display (HMD) as an individual navigation device [5]. These systems reproduce the object's depth with binocular vision, which may not give the observer the correct sense of depth.

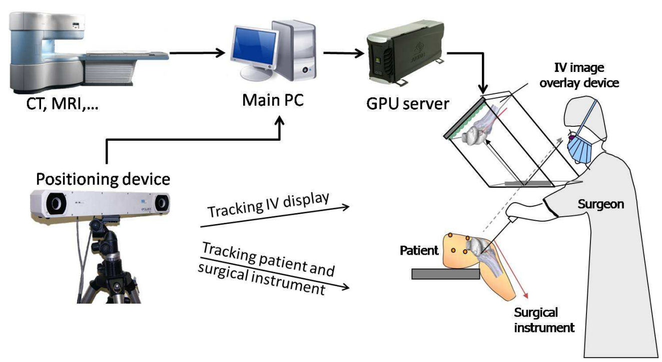


Fig. 1. Configuration of IV image guided system

Also, motion parallax can only be reproduced using external tracking devices that come along with many physical restrictions. To solve this problem, Liao et al. [7] developed an image overlay system based on the Integral Videography (IV) visualization method that allowed auto-stereoscopic images to be viewed with the naked eye. However, the rendering speed depends greatly on the data size and some visual effects can not be achieved. Alternatively, a computer graphics (CG) based IV rendering method, the pixel distribution[8] was introduced.This method takes the advantage of CG technologies but still it faces a rendering speed problem because of the large amount of computation required. In this paper, we aim to develop a fast, CG based auto-stereoscopic imaging algorithm with interactive and intuitive vision of the surgical site and to create a compact, flexible system in order to remove the potential physical restrictions which may affect user's performance.

2 Materials and methods

2.1 3D auto-stereoscopic image guided surgical navigation system

The system consists of an image acquisition device (CT or MRI), an optical position tracking device, a main PC, and an IV display (Fig.1). Three-dimensional (3D) data of the anatomy from the image acquisition device is sent to the main PC through RS232C, where the data will then be segmented and reconstructed into 3D surface models used to render IV images. The rendering task is done by an external multi-GPU server connected to the main PC through a DVI link. During the operation, positions and orientations of optical markers planted on the surgical tool are tracked by the positioning device and are updated every 30 milliseconds. This setting allows surgeons to know exactly the positional relationship between the patient and the surgical tool.

The IV display consists of a micro convex lens array and high-density flat LCD display. This display is placed at the focal plane of the lens array so that

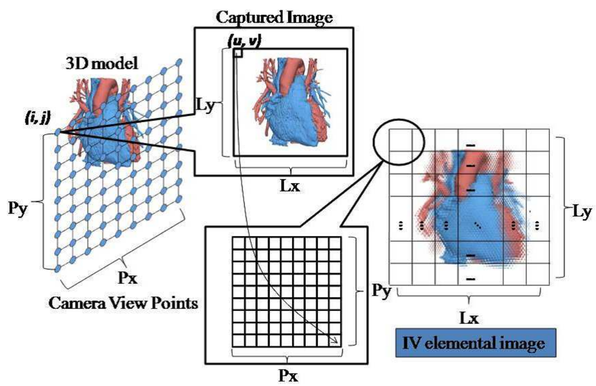


Fig. 2. Pixel distribution algorithm

light rays from corresponding pixels will converge at a single dot in real space. The first principle of IV, the Integral Photography or IP was developed by Lippmann [6] where 3D images are recorded and reproduced using a film sheet. Using an LCD, we now can display animated images instead of only still images.

To create an IV image, we need a set of multi-view images which can be acquired using the conventional geometric based surface rendering. Each pixel of the viewpoint image is distributed to the pixel on the display that lies on the same light ray. This method uses CG surface models as the input and produces high-quality images with better visual effect. Also, peripheral devices such as surgical tool can be visualized in IV image as a simple CG model.

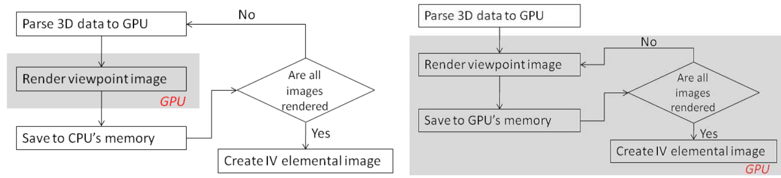
2.2 IV image rendering

Automatic rendering of IV images includes rendering viewpoint images and performing pixel distribution (Fig. 2). Supposed that there are $L_X \times L_Y$ micro lenses in the lens array and each micro lens covers approximately $P_X \times P_Y$ pixels on the background image. We will need to capture $P_X \times P_Y$ images at $P_X \times P_Y$ viewpoints corresponding to the number of pixels covered by one lens. Let's denote the pixel $(u, v)(u = 0..L_X - 1, v = 0..L_Y - 1)$ of the image at viewpoint $(i, j)(i = 0..P_X - 1, j = 0..P_Y - 1)$ as $P_{(i,j,u,v)}$ and the pixel (i, j) behind the lens(u, v) as $L_{(u,v,i,j)}$. The following relation holds,

$$P_{(i,j,u,v)} = L_{(u,v,P_X-i,P_Y-j)}. \tag{1}$$

2.3 High speed rendering with GPU

Conventionally, rendering viewpoint images and calculating pixel distribution are done by GPU and CPU respectively. In this paper, we proposed a GPU-



(a) (b)

Fig. 3. IV Rendering work's flow using (a) conventional method, (b) proposed method

based algorithm which performs all of the computations on GPU only. The main idea is, viewpoint images are rendered and saved directly to the GPU's memory instead of the CPU's memory so that pixel distribution can be done on the GPU afterward.

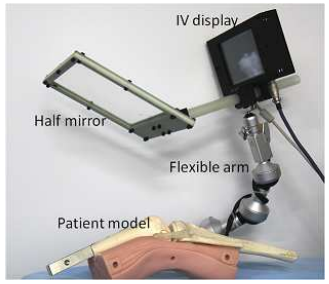
Images from $P_X \times P_Y$ viewpoints are rendered all together at the same time by dividing the screen into $P_X \times P_Y$ viewports. Each viewpoint image has a size of $L_X \times L_Y$ pixels (it takes over a $L_X \times L_Y$ portion of the screen). Next, we save these viewpoint images to GPU's memory by using the Frame Buffer Object (FBO). This is actually a render-to-texture technique that allows images from a frame buffer to be sent to a texture buffer rather than a display device.

The next thing we have to take care of is how to use this texture buffer. Conventionally texture buffer is to be mapped onto a surface and that may be the only way to deal with the texture buffer. So in the next step we draw a plain rectangle which has the same size with the display screen and map the texture onto it. The problem that comes out is: how can we perform pixel distribution on that texture?

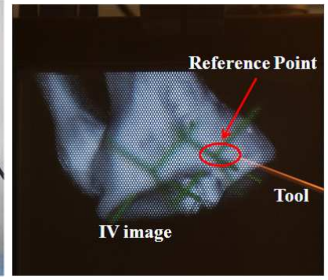
Pixel distribution is performed when dealing with a single fragment (a square portion of the screen which is a potential pixel). To deal with this, we use the C for Graphics (or Cg) [9] language. Cg is a general purposed GPU programming language that gives programmers the access to the graphics pipeline and let them reprogram the fragment process of the GPU. Each fragment input to the Cg fragment program has a texture coordinate (which corresponds with its actual position on the screen in this case). Let's note that the whole screen would has a size of $(P_X \times L_X) \times (P_Y \times L_Y)$ pixels, we can denote the position of the fragment as $(u \times P_X + u, v \times P_Y + j)$ where $(i = 0..P_X - 1, j = 0..P_Y - 1)$ and $(u = 0..L_X - 1, v = 0..L_Y - 1)$.

The four parameter (i, j, u, v) can be computed easily using the texture coordinate of the fragment. They indicates that the fragment is at the pixel (i, j) behind the lens number (u, v). Using equation 1 we can compute which pixel of the texture buffer should come to that position. The rest is very simple, we change the input texture coordinate to the newly computed one and output the fragment with new texture coordinate to the rendering machine.

The main benefits of our algorithm include data transfer between GPU and CPU becomes unnecessary and computation of pixel distribution is accelerated



(a)



(b)

Fig. 4. IV image overlay system: oveview of the system(a), accuracy eveluation by touching the reference points (b).

with the GPU's architecture (Fig. 3). IV images can therefore be rendered in much higher speed than in conventional method.

3 Experiments and results

3.1 Image overlay

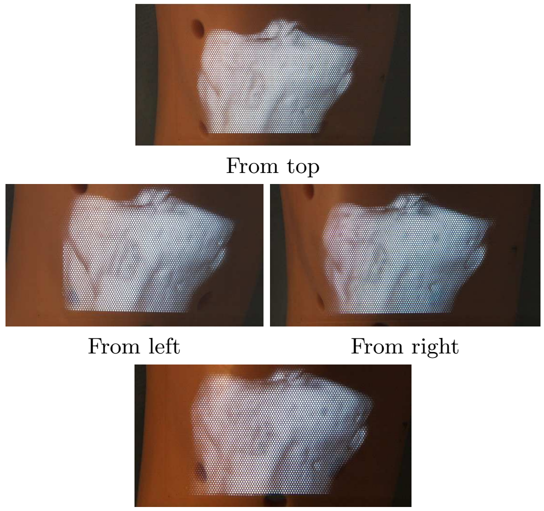
The IV overlay device consisted of $20 \times 20cm$ half-silvered mirror which was placed at 45° and $25[cm]$in front of the IV display so that the IV image would appear almost perpendicularly to the observer (fig.4 (a)). The specifications of the IV display system was as follow:

Table 1. Specifications of the IV display system

Display	6.4 inch display, 203 dpi(XGA)
Lens array	Fly eye lens array
Horizontal lens pitch	1.001 mm (Covers approximately 8 pixels)
Vertical lens pitch	0.876 mm (Covers approximately 7 pixels)

The overlay device was attached to a flexible arm and can be removed easily from the surgical site. Also, the surgeons can eventually translate and rotate the overlay device in order to have different views of the object, or even different parts of the patient as long as the data is prepared. This configuration gives the surgeons a wider and more flexible visual access to the surgical area compared to the conventional systems.

In normal 2D based image guided systems, registration is performed by calculating the transformation matrix between the computer space where the model



From top

From left

From right

From bottom

Fig. 5. IV image overlaid with patient model.

exists and the real space where the patient exists (the ${}^{Patient}_{Model}T$ matrix). In our system, beside of ${}^{Patient}_{Tracker}T$, we also need to compute the ${}^{IV}_{Tracker}T$ matrix which registers the IV image to the surface model. Transformation matrix between the IV image and the patient is calculated as follow.

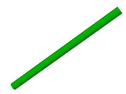
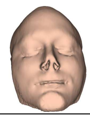
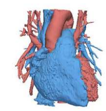
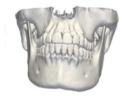
$$ {}^{IV}_{Patient}T = {}^{IV}_{Tracker}T \times {}^{Patient}_{Tracker}T^{-1} \tag{2} $$

Both ${}^{Patient}_{Model}T$ and ${}^{IV}_{Model}T$ were computed using the fiducial points based registration method [10, 11]. We evaluated the accuracy of this overlay system by touching a set of two reference points with the marked tool to record their positions in real space. Then we compared them to the actual positions on the bone(Fig. 4(b)). Our overlay system has a positional error of $1.7 \pm 0.6mm$. The overlaid IV images taken from 4 directions of a human tibia are shown in Fig. 5.

3.2 Evaluation of rendering speed

We used a number of surface models with various complexity to evaluate the improvement in rendering speed. The anatomy models were created from CT data while the model of the surgical tool was simply a cylinder tube. The number of polygon faces of these models varied from 8 (the surgical tool model) to 486674 (the human skull model). We used an Intel Core 2 6600@2.4GHz CPU with 4GB of memory and a NVIDIA Quadro Plex model 4 GPU server which contains two Quadro FX 5600 GPUs accelerated by Scalable Link Interface (SLI) technique.

Table 2. Rendering time for 1 frame (ms)

Model	Tool	Head	Bone	Heart	Skull
					
Number of cells	8	22915	97119	169829	486674
Conventional method	52 ± 5	412 ± 61	692 ± 5	1158 ± 38	1395 ± 9
Proposed method	6 ± 0	45 ± 0	85 ± 1	321 ± 6	408 ± 0

For the lens array and the display discribed in Table 1, $8 \times 7 = 56$ views were required for a single IV image. Because the display's size was 1024×768 pixels, each viewpoint image had dimensions of 128×110 pixels.

The rendering speed for one IV image frame is shown in Table 2. With the simplest model, it took 52 ± 5 milliseconds to render an IV frame of XGA size using the conventional CPU based method. In contrast, the proposed GPU based method achieved a rendering speed of 6 ± 0 milliseconds for one IV frame, meaning an improvement of 867% in rendering speed compared to the CPU based method. With the most complicated model, this improvement was 304%.

4 Discussions

The positional error of the overlay system is still too large for the system to be usable in clinical cases (1.7 mm). This is because performing a two-step registration resulted in a doubled error. However, this error can be reduced with direct registration between the IV image and the patient and other calibration methods which we are planning to do in the future.

With most of the models, it took less than 0.1 ms to render one frame, meaning a frame rate of more than 10 frames per second (fps). This rendering speed allows surgeons to smoothly perform interactions with the IV image in almost real time. This feature is particularly useful in surgical navigation systems where real-time vision of the surgical site is consistently required. In particular, the fact that IV images of the surgical tool can be rendered at extremely high speed (6 ms) indicates that every motion the surgeon makes can be reported on the IV display with almost no time delay.

Decrease in the speedup can be noticed as the model becomes complex. The reason is simple, with less complex model, rendering a single viewpoint image (which is executed in exactly the same way in both conventional and proposed method) costs less than the other tasks, while with complex model, it becomes the main time consuming task. One solution for dealing with this problem is to combine our algorithm with other IV rendering techniques such as image-based rendering or volume cay-casting.

5 Conclusion

We have built a high-speed IV image overlay system for surgical navigation. By implementing the IV rendering algorithm on the graphics hardware, we have greatly improved the frame rate of IV images that allows surgeons to interacts with the image in real-time. With better registration method, we hope to reduce the positional error to under 1mm to make our system usable in clinical case.

Acknowledgment

This study was supported in part by the Special Coordination Funds for Promoting Science and Technology commissioned by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in Japan, Grant for Industrial Technology Research (07C46050) of New Energy and Industrial Technology Development Organization, Japan, and the Japan Society for the Promotion of Science (219746).

References

1. IM. Germane, H. Villalobos, A. Silvers, KD. Post: Clinical use of the optical digitizer for intracranial neuronavigation. Neurosurgery, Vol.45, No.2, pp.261-269 (1999)
2. Plaweski S, Cazal J, Rosell P, Merloz P: Anterior cruciate ligament reconstruction using navigation: a comparative study on 60 patients. The American Journal of Sports Medicine 34:542-552 (2006)
3. Azar, F.S., Perrin, N., Khamene, A., Vogt, S., Sauer, F.: User performance analysis. of different image-based navigation systems for needle placement procedures. Medical Image Computing and Computer-Assisted Intervention, pp.373-380 (2006)
4. M. Blackwell, C. Nikou, Anthony M.DiGioia, T. Kanade: An Image Overlay system for medical data visualization. Medical Image Analysis, Vol.4, pp.67-72 (2000)
5. Joerg Traub, Tobias Sielhorst, Sandro-Michael Heining, and Nassir Navab: Advanced Display and Visualization Concepts for Image Guided Surgery. Journal of Display Technology, Vol. 4, Issue 4, pp. 483-490 (2008)
6. M. G. Lippmann: Epreuves reversibles donnant la sensation du relief. J. de Phys., Vol.7, pp.821-825 (1908)
7. H. Liao, S. Nakajima, M. Iwahara, N. Hata, I. Sakuma, T. Dohi: Real-time 3D Image-guided Navigation system based on Integral Videography. Proceedings of SPIE, Vol.4615, pp.36-44 (2002)
8. Hongen Liao, Keisuke Nomura, Takeyoshi Dohi: Autostereoscopic Integral Photography Imaging using Pixel Distribution of Computer Graphics Generated Image; ACM SIGGRAPH 2005, CD-ROM, Los Angeles, USA (2005)
9. Randima Fernando, Mark J. Kilgard: The Cg Tutorial. (2003)
10. C. R. Maurer, J. M. Fitzpatrick, M. Y. Wang, R. L. Galloway, R. J. Maciunas, G. S. Allen: Registration of head volume images using implantable fiducial markers. IEEE Trans. on Med. Imag., Vol.16, pp.447-462 (1997)
11. K. S. Arun, T. S. Huang, and S. D. Blostern: Least-square fitting of two 3-D point set. IEEE Trans. Pattern Anal. Machine Intell., Vol.PAMI-9, pp.698-700 (1987)

Fast non-rigid registration of medical image data for image guided surgery

S.K. Shah[1], R.S. Rowland[2], and R.J. Lapeer[1]

[1] School of Computing Sciences, University of East Anglia, Norwich, UK
[2] RMR Systems Ltd., Ipswich IP1 1TT, UK

Abstract. With the increasing importance of merging multiple datasets of different imaging modalities for image guided surgery applications, the speed by which such datasets can be registered is crucial. This paper presents a novel approach based on fast Radial Basis Function (RBF) evaluation using the biharmonic spline (BHS) and thin Plate spline (TPS) models for non-linear registration. The new algorithm is evaluated both in standard software and hardware assisted versions. A number of experiments on medical image data are presented, illustrating substantial speedups compared to a 'brute-force' implementation of a radial basis function based registration algorithm. The global transformation accuracy of these techniques has been evaluated by using two sets of anatomical landmarks, one set for calculating the model parameters, and the other set for assessing registration accuracy. Finally, we demonstrate that the technique yields sub-second performance and target registration errors of about 1.2mm on the intra-subject registration of CT and MRI image datasets obtained from the Vanderbilt database, when implemented on the GPU of a standard PC with high-end video-adapter card.

1 Introduction

A significant proportion of multi-modal registration problems require non-rigid models. A number of publications show that non-rigid registration does indeed give better results than rigid registration when deformable tissue is involved. Rueckert et al. [1] demonstrated the superiority of free-form deformations based on B-splines when compared to rigid and affine transformations applied to MRI breast images. Fei et al. [2] showed that non-rigid warping consistently outclassed rigid registration for prostate and pelvic MRI data. Non-rigid registration of two medical images involves the identification (either manually or partially automated) of correspondences between images, followed by a transformation which maps one image onto the other. Four typical categories are distinguished: point-based, surface-based, intensity-based and model-based. Point-based registration involves the identification of a set of corresponding points in the images being registered, which may be manual or semi-automated. Maintz and Viergever [3] distinguished between intrinsic and extrinsic registration, based on respectively, the absence or presence of fiducial markers attached (internally or externally) to the patient. The main drawback of an intrinsic method is the time-consuming

identification of landmarks. Though fiducial markers (extrinsic method) are easier to identify in the image, a range of different issues arise, such as pre-operative preparation, patient discomfort and displacement or replacement errors. The major advantage of point-based methods in general is their fast computation time. Surface based registration methods require the extraction of equivalent surfaces which implies segmentation of the raw data. This in turn implies that the registration is dependent on the accuracy of the surface segmentation. Generally, surface-based methods have been confined to neuro-imaging and orthopaedic applications, image-guided surgery and atlas-based brain mapping and segmentation. In the field of Endoscopic Sinus Surgery, Hardy et al. [4] report that surface registration is superior, both in terms of speed and accuracy, to point-based registration methods (with or without fiducials). However, other studies show that surface-based methods are inferior to intensity-based methods [5]. Intensity-based methods require less pre-processing than the previous two techniques and operate directly on the grey values in the images. The most prevalent intensity-based methods use mutual information (MI) as a measurement of image similarity. Maes et al. [6] and Pluim et al. [7] provide good surveys of MI based registration methods. The crucial characteristic of the optimisation process is the convergence to a global rather than a local maximum which can be time consuming. Finally, model-based registration builds a model of the deformation field based on the physical characteristics of the organ or tissue of interest. Typical modelling techniques used are elastic modelling, the finite element method (FEM) and fluid dynamics.
Previous work to improve the speed of non-rigid medical image registration has primarily targeted intensity based methods and a number of research groups have employed hardware solutions with or without parallel architectures or algorithms. More recently, Levin et al. [8] proposed a method for accelerating point-based non-rigid registration by using the fast tri-linear interpolation capability of modern graphics cards. Their implementation evaluates a thin-plate spline (TPS) warp at discrete points on a configurable sized grid which overlays each image data slice. Interpolation by the graphics card is used to estimate the intensity values of the voxels within each grid cell. They showed that their grid-based warp reduced registration time from 148.2s to 1.63s for a 512x512x173 dataset using 92 user defined landmarks. Finally, a recent comparison of acceleration techniques for non-rigid medical image registration can be found in Klein et al. [9]. In this paper, we present a novel method to accelerate the evaluation of the transformation function for non-rigid medical image registration, primarily aimed at (though not restricted to) image guided surgery (IGS) applications, using a point-based method.

2 Methodology

2.1 Choice of Radial Basis Function

The Radial Basis Function (RBF) method is one of the most widely used techniques for approximating or interpolating scattered data in multiple dimensions.

When considering interpolation, the aim is to approximate a real-valued function $f(\mathbf{x})$ for which we have a finite set of values $f = (f_1, \ldots, f_N)$ at the distinct points $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \subset \mathbf{R}^d$. In this situation, we can choose a RBF, $s(\mathbf{x})$, to represent such an approximation, which has the general form :

$$s(\mathbf{x}) = p(\mathbf{x}) + \sum_{i=1}^{N} \lambda_i \phi\left(\|\mathbf{x} - \mathbf{x}_i\|\right), \qquad \mathbf{x} \in \mathbf{R}^d \tag{1}$$

where $p(\mathbf{x})$ is a polynomial, λ_i is a real-valued weight[3], ϕ is the (radial) basis function and $\|\mathbf{x} - \mathbf{x}_i\| = r$ is the Euclidean distance between $\mathbf{x}$ and $\mathbf{x}_i$. So, a RBF could be described as a weighted sum of a radially-symmetric basis function, augmented by a polynomial term. The basis function ϕ can take several forms. Rohr [10] shows that the biharmonic spline (BHS): $\phi(r) = r$, minimises a bending energy potential of order two in three dimensional space. As we aim to warp three-dimensional image data, the BHS is therefore the preferred choice. However, the BHS is only C^0 continuous in its centre, whereas the TPS, $\phi(r) = r^2 \log r$, is C^1 continuous in its centre. Therefore we will also consider the TPS.

2.2 Fast RBF

Let us rewrite Equation 1 by omitting the linear polynomial part for sake of clarity, and extend to 3D for evaluation of $i = 1 \ldots m$ targets (voxels) represented by the target (voxel) vector $\mathbf{x}_i$, after having found the spline parameters λ_j for $j = 1 \ldots n$ landmarks represented by the source (landmark) vector $\mathbf{y}_j$:

$$s(\mathbf{x}_i) = \sum_{j=0}^{n} \lambda(\mathbf{y}_j)\phi(\|\mathbf{x}_i - \mathbf{y}_j\|), \qquad i = 0, 1, \ldots, m. \tag{2}$$

Livne and Wright [11] describe a new method for fast multilevel evaluation of RBF expansions. The basic idea of the fast RBF method is that a smooth RBF, ϕ, can be represented accurately on a regular coarse grid, of fewer nodes than the full voxel set and that the expensive summation in Equation 2 need be performed only at these nodes while the remaining voxel values can eventually be determined using a less expensive formulation based on the values calculated for the surrounding nodes. Unlike the grid based approach by Levin et al. [8], it is the RBF coefficients that are interpolated within the grid and not the intensity values of the voxels. Figure 1(a) illustrates the main principle of the method by encapsulating source and target points in separate grids of size H. This results in a two stage process conversion of the RBF in Equation 2. The first

[3] The λ weights are determined in the 'calculation' step using a least mean squares approach. This step is followed by the 'evaluation' step which applies the RBF to (usually) all voxels. The latter step is much more time-consuming than the former, hence it is this step which we aim to optimise.

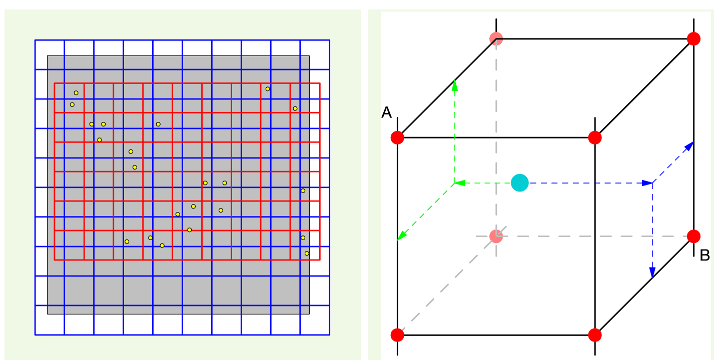
stage replaces the original **source** points with their corresponding grid points by using a centered p-th order tensor product interpolation:

$$\phi(\|\mathbf{x}_i - \mathbf{y}_j\|) = \sum_{j:J_k \epsilon \sigma_j^{(k)}} \omega_{jJ_3}\omega_{jJ_2}\omega_{jJ_1}\phi(\|\mathbf{x}_i - \mathbf{Y}_{(J_1,J_2,J_3)}\|) \tag{3}$$

where $j = 0, 1, \ldots, n$ and for dimension $k = 1, 2, 3$:
$\sigma_j^{(k)} := \left\{ J_k : |Y_{J_k}^{(k)} - y_j^{(k)}| < pH/2 \right\}$, where ω_{jJ_k} are the new centered pth-order interpolation weights from the coarse centres $Y_{J_k}^{(k)}$ to the landmark positions $y_j^{(k)}$. The second stage replaces the original **target** points with their corresponding grid points using the same approach:

$$\phi(\|\mathbf{x}_i - \mathbf{Y}_{\mathbf{J}}\|) = \sum_{I_k \epsilon \bar{\sigma}_i^{(k)}} \bar{\omega}_{iI_3}\bar{\omega}_{iI_2}\bar{\omega}_{iI_1}\phi(\|\mathbf{X}_{(I_1,I_2,I_3)} - \mathbf{Y}_{\mathbf{J}}\|) \tag{4}$$

where $i = 0, 1, \ldots, m$, $\mathbf{J} = (J_1, J_2, J_3)$, and for dimension $k = 1, 2, 3$:
$\bar{\sigma}_i^{(k)} := \left\{ I_k : |X_{I_k}^{(k)} - x_i^{(k)}| < pH/2 \right\}$, where $\bar{\omega}_{iI_k}$ are the centered pth-order interpolation weights from the coarse evaluation point $X_{I_k}^{(k)}$ to the level h (original image grid size) evaluation point $x_i^{(k)}$. The procedure used to distribute the known RBF coefficients $\lambda(\mathbf{y}_j)$ at each landmark position to the surrounding nodes of grid $\mathbf{Y}$ is called *anterpolation* - see Figure 1(b). In depth coverage of the method in 1D and 2D can be found in [11] and its 3D extension in [12].



(a) (b)

Fig. 1. (a) 2D example of the $\mathbf{X}$ 'target' grid (outer/blue) and $\mathbf{Y}$ 'source' grid (inner/red) with (yellow) landmarks for "Fast RBF". Both grids have size H squares. (b) Anterpolation in 3D: a single (centre/cyan) landmark within grid $\mathbf{Y}$, showing the (vector) elements of interpolation weights for nodes A (light/green vectors) and B (dark/blue vectors).

2.3 Choice of performance metric

We adopt the following two performance metrics:
Target Registration Error (TRE): The TRE is the RMS error between the homologous validation landmarks after registration. To help the evaluation of global accuracy of registration we developed a set of well defined validation anatomical landmarks which are distributed across the dataset.
Normalized mutual Information (NMI): As the NMI metric (Studholme et al. [13]) is suited to both mono-modal and multi-modal scenarios, we use this metric for image similarity measurement. This metric is overlap invariant, which means that it does not depend on the degree of overlap of the two images and has an optimal value of 2.0 and a minimum value of 1.0.

3 Experiments and Results

We report performance during the evaluation stage of five different implementations which are: (1) Brute force (non-optimised) RBF; (2) Brute force (non-optimised) RBF with hardware acceleration; (3) Fast RBF; (4) Fast RBF with hardware acceleration; (5) Grid approach by Levin et al. [8].
In [14] we presented a series of experiments on synthetic data. We showed that we could achieve sub-second performance on warping a checkered cube of 256^3 voxels using the hardware implementation on the GPU of the fast RBF algorithm. The accuracy of the warp remained at a respectable level of 86% as compared to the brute-force software based algorithm used as the gold standard.
In this paper, we present a set of experiments on real image data as obtained from the Vanderbilt Database, also known as the Retrospective Image Registration Evaluation (RIRE) project [15]. RIRE is a project which uses a rigid registration method based on fiducial markers as a gold standard. The image database is free for use to any research team which wishes to test and compare their (rigid) registration technique. This does not exclude its use to test non-rigid registration methods, however comparison of results with the site's gold standard should therefore be interpreted with caution. As we are using real image data this time round, we have no ground truth when using a metric such as NMI. Therefore, we first warp each of the datasets (CT to MRI) and then warp the deformed dataset back to the original. This allows us to compare the NMI's of a twice deformed dataset with its original, using the latter as the ground truth[4].
To evaluate the speed-optimised algorithms which use hardware acceleration, i.e. (2),(4) and (5), in terms of accuracy, the brute force algorithm (1) is considered as the gold standard. This is because current GPU's, despite being significantly faster than CPU's, only have 32 bits for floating point representation, whereas CPU's have 64 bits, which affects the accuracy of the warp when measured using the NMI (see later). The MR and CT datasets of the Vanderbilt database [15] were used (P001-P006;P101-P102) and resampled to 256^3 with slice thicknesses of 1mm. We tested all algorithms using 8 and 16 landmarks respectively, for "training" (to fit the spline). For both these experiments 8 validation landmarks (for testing) were used - Table 1.

[4] Note that the TRE error is evaluated for a forward warp only.

8 landmarks	Eval. Time in sec.	NMI	%NMI	TRE in mm.
Brute force S/W TPS	28.096(2.020)	1.556(0.052)	100.0	1.59(0.23)
Brute force H/W TPS	0.509(0.070)	1.301(0.027)	83.6	1.59(0.23)
Fast RBF S/W TPS	13.010(0.294)	1.555(0.053)	99.9	1.59(0.23)
Fast RBF H/W TPS	0.443(0.016)	1.369(0.036)	88.0	1.59(0.23)
Levin TPS (H/W)	0.412(0.025)	1.349(0.026)	86.7	1.59(0.23)
Brute force S/W BHS	21.438(1.020)	1.562(0.053)	100.0	1.56(0.26)
Brute force H/W BHS	0.485(0.026)	1.300(0.027)	83.2	1.56(0.26)
Fast RBF S/W BHS	13.024(0.309)	1.560(0.055)	99.9	1.56(0.26)
Fast RBF H/W BHS	0.438(0.009)	1.369(0.037)	87.6	1.56(0.26)
Levin BHS (H/W)	0.371(0.015)	1.347(0.026)	86.2	1.56(0.26)
16 landmarks	Eval. Time in sec.	NMI	%NMI	TRE in mm.
Brute force S/W TPS	42.559(4.345)	1.587(0.015)	100	1.22(0.19)
Brute force H/W TPS	0.595(0.034)	1.318(0.015)	83.0	1.22(0.19)
Fast RBF S/W TPS	12.880(0.234)	1.577(0.021)	99.4	1.22(0.19)
Fast RBF H/W TPS	0.446(0.007)	1.388(0.007)	87.5	1.22(0.19)
Levin TPS (H/W)	0.504(0.068)	1.356(0.034)	85.4	1.22(0.19)
Brute force S/W BHS	31.644(3.687)	1.592(0.013)	100	1.21(0.17)
Brute force H/W BHS	0.614(0.050)	1.319(0.015)	82.9	1.21(0.17)
Fast RBF S/W BHS	12.831(0.046)	1.587(0.015)	99.7	1.21(0.17)
Fast RBF H/W BHS	0.465(0.012)	1.388(0.007)	87.2	1.21(0.17)
Levin BHS (H/W)	0.448(0.020)	1.360(0.030)	85.4	1.21(0.17)

Table 1. Results after applying TPS and BHS basis functions for non-rigid registration of the CT and MR-T1&T2 RIRE data. In the upper table 8 landmarks were used for training, whilst 16 landmarks were used for the experiments as shown in the lower table. For both sets of experiments, 8 landmarks were used for validation and all tests were run over 8 subjects. Values are averages with standard deviation in parentheses. The second column shows the evaluation time of the RBF in seconds. The third column shows the NMI after warping forwards and backwards. The next column shows the %NMI as compared to the Brute-Force Software used as the golden standard. The fifth and final column shows the TRE in mm. which is evaluated on the validation landmarks (forward warp only).

4 Discussion

Let us first have a look at the speed of the different algorithms. We notice that the Fast RBF in software (S/W) is more than twice as fast for the TPS basis function (and slightly less so for the BHS) than the brute force software implementation for 8 landmarks (Table 1 - upper half). Looking at the same algorithms in Table 1 - lower half, we see that the speedup factor has increased to 3-4. If we consider the hardware accelerated (H/W) algorithms we notice that the best performances (< 0.5s) are displayed by the Fast RBF algorithm and the Grid-based method by Levin et al. [8]. However, whilst the latter method performs substantially worse when more landmarks are used, the fast RBF method is significantly less affected. This is in line with what we already saw for the software implementation. Indeed, the evaluation speed of the fast RBF method is significantly less dependent on the number of landmarks used than for competing methods. Looking next at the accuracy using the NMI metric, we see virtually no loss in accuracy for the fast RBF method in software as compared to the golden standard (brute force software) and when implemented in hardware, its accuracy is better than the other tested hardware implementations. Finally, the TRE, which is independent of the speed optimisation method used is acceptable for IGS applications at an average value of around 1.2mm (16 training landmarks) as it is of the same order of magnitude as registration errors of commercial surgical navigation systems. There are no significant differences in accuracy between the TPS and BHS functions ($p >> 0.10$), save for the latter being faster for the brute force software implementation, because of the additional *log* function to be evaluated for the TPS. The fast RBF method, either implemented in hardware or in software, is not affected by this.

5 Conclusion

We have presented the Fast RBF method for fast non-rigid registration of medical imaging data using anatomical landmarks. The algorithm when implemented in hardware yields sub-second performance on a standard PC with high-end video adapter card. The warp time of the Fast RBF algorithm, irrespectively of being implemented in software or in hardware, is significantly less susceptible to the number of landmarks used as compared to the tested competing methods. Considering that more accurately placed landmarks improve accuracy implies that this algorithm is favourable for applications where both speed and accuracy are of importance, such as in IGS. Future work will cover experiments on medical image data with higher degrees of non-rigid distortion, for example MR images of the brain, before and after brain shift.

References

1. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J.: Nonrigid registration using free-form deformations: Application to breast MR images. IEEE Transactions on Medical Imaging **18**(8) (1999) 712–721
2. Fei, B.W., Kemper, C., Wilson, D.L.: A comparative study of warping and rigid body registration for the prostate and pelvic MR volumes. Computerized Medical Imaging and Graphics **27**(4) (2003) 267–281
3. Maintz, J.B.A., Viergever, M.A.: A survey of medical image registration. Medical Image Analysis **2**(1) (1998) 1–36
4. Hardy, S.M., Melroy, C., White, D.R., Dubin, M., Senior, B.: A comparison of computer-aided surgery registration methods for endoscopic sinus surgery. Am J Rhinol **20**(1) (2006) 48–52
5. West, J., Fitzpatrick, J.M., Wang, M.Y., Dawant, B.M., Maurer, C.R., Kessler, R.M., Maciunas, R.J.: Retrospective intermodality registration techniques for images of the head: surface-based versus volume-based. IEEE Trans Med Imaging **18**(2) (Feb 1999) 144–150
6. Maes, F., Vandermeulen, D., Suetens, P.: Medical image registration using mutual information. Proceedings of the IEEE **91**(10) (2003) 1699–1722

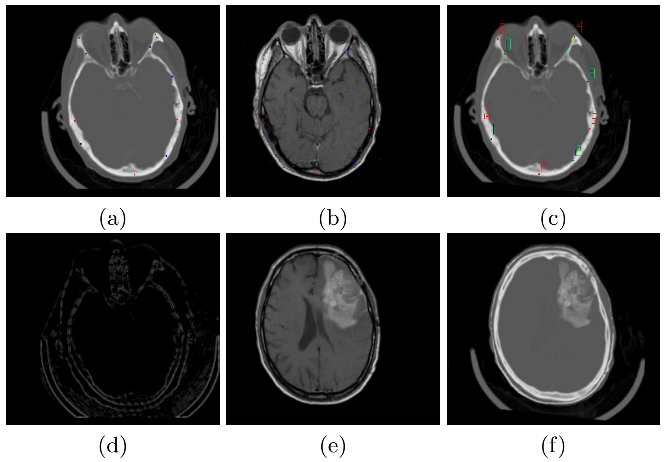


Fig. 2. Selected transverse slices from the full resolution CT-to-MRI datasets from the Vanderbilt database, before (a,b) and after (c) the registration experiment. The second rows shows the absolute difference image (gamma-corrected - $\gamma = 1.6$ - for clarity) of the original CT image after warping forwards and back to its ground truth using the fast RBF software (d). Next, another MRI slice with visible tumor is shown, before (e) and after fusing it with a registered CT image (f).

7. Pluim, J.P.W., Maintz, J.B.A., Viergever, M.A.: Mutual-information-based registration of medical images: a survey. IEEE Trans Med Imaging **22**(8) (Aug 2003) 986–1004
8. Levin, D., Dey, D., Slomka, P.: Efficient 3D nonlinear warping of computed tomography: two high-performance implementations using OpenGL. In Galloway, Jr., R., Cleary, K., eds.: Medical Imaging 2005: Visualization, Image-Guided Procedures, and Display. Proceedings of the SPIE. Volume 5744. (2005) 34–42
9. Klein, S., Staring, M., Pluim, J.: A comparison of acceleration techniques for nonrigid medical image registration. In: International Workshop on Biomedical Image Registration. Volume 4057 of Lecture Notes in Computer Science. (2006) 151–159
10. Rohr, K.: Landmark-Based Image Analysis Using Geometric and Intensity Models. Kluwer Academic Publishers (2001)
11. Livne, O.E., Wright, G.B.: Fast multilevel evaluation of smooth radial basis function expansions. Electronic Transactions on Numerical Analysis **23** (2006) 263–287
12. Rowland, R.: Fast Registration of Medical Imaging Data using Optimised Radial Basis Functions. PhD thesis, University of East Anglia (2007)
13. Studholme, C., Hill, D.L.G., Hawkes, D.J.: An overlap invariant entropy measure of 3D medical image alignment. Pattern Recognition **32** (1999) 71–86
14. Rowland, R., Lapeer, R.: Fast non-rigid registration of medical images: comparing the thin-plate and biharmonic splines. In: Medical Image Understanding and Analysis, MIUA 2008, Dundee. (2008) 246–250
15. Fitzpatrick, J.: Retrospective image registration and evaluation project (RIRE) – http://www.insight-journal.org/rire/index.php (1997)

Statistical Deformation Model Driven Atlas Registration with Stochastic Sub-sampling

Wenzhe Shi[1], Lin Mei[1], Daniel Rueckert[2], Ara Darzi[1], and Philip Edwards[1,2]

[1] Dept. of Biosurgery and Surgical Technology, Imperial College London, UK
[2] Dept. of Computing, Imperial College London, UK
l.mei,eddie.edwards@imperial.ac.uk *

Abstract. For augmented reality guidance of robotic prostatectomy, the first step is to build a model of the patient by segmentation of preoperative images. The aim of this work is to apply statistical shape modelling to automate this process. Statistical deformation models (SDMs) have been proposed as a means of shape analysis using the node points of non-rigid registration. This has the advantage that correspondence comes from the registration process and is established automatically. Such a SDM is created for the lower abdomen with the aim of segmentation of the pelvis, prostate and rectum for image guidance of robotic prostatectomy. For the individual registrations in the training set, an initial registration using manual landmarks was required for accurate alignment to the atlas. We propose that future registrations can be guided by this model by restricting the non-rigid registration to the principal modes of the SDM.
Several technical difficulties arose. Firstly, the SDM modes represent movement of all of the b-spline node points, which leads to inefficient calculation of the cost function gradients. This is overcome by projecting the gradient at the node points onto the principal modes of the SDM. Rather than limiting the registration to being along the principal modes of the SDM, a cost function for deviation from the space of the SDM can be applied. This allows greater freedom of node point motion but encourages realistic deformations. Efficiency can be further improved by stochastic sub-sampling in a region of interest dependent on the influence each node point position has on a given mode. During the iteration, modes that have little influence on the cost function gradient are ignored, but can be reinstated with a given probability. The result is an atlas registration that is 8 times faster than ordinary non-rigid registration and does not require initial manual landmark identification.

1 Introduction

Statistical shape models are a popoular class of techniques that encapsulate the variation of anatomy into a reduced dimensional set of principal modes [1]. Such

* We would like to thank Tyco Healthcare for funding Lin Mei's PhD studentship. We are also grateful to many other members of the Department of Computing and the Department of Biosurgery and Surgical Technology at Imperial College.

models are built using a training set consisting of a large number of examples with corresponding landmarks, usually on the surface of the target organ. Shape models have been proposed for segmentation of the lower abdomen [2]. In a statistical deformation model (SDM) the correspoinding points are the control points of a b-spline grid that describe a free-form deformation of a template scan or atlas to images of the individuals in the training set [3]. This has the advantage that the correspondence is automatically generated by the process of non-rigid registration. For an active deformation model, we can then restrict the deformation from the atlas to a new individual using the principal modes of variation of the SDM.

In this paper, we proposed and implemented a detailed scheme of a SDM-based methodology for segmentation and surface generation from magnetic resonance imaging (MRI) of the lower abdomen. There are two main stages. Firstly the SMD is created from training set which invoves registration of a segmented template scan to many example images semi-automatically. The second stage is to use the statistical deformation model to do a fully-automatic registration to new scans. We can use the resulting registration to propagate the individual surface model from atlas. The SDM is to be used to reduce the search space of non-rigid registration. Two different methods to use the SDM to guide the non-rigid registration process were implemented and the accuracy and complexity of the methods were compared. The result is a complete scheme of fully automatic segmentation of MRI images and generation of tissue surface models of the lower abdomen. The eventual aim is to use these segmentations for augmented reality guidance of robotic prostatectomy. Comparison with non-rigid registration methods and of different data sets are given. Advantages and disadvantages of the methods are analyzed.

2 Methods

2.1 Initial Hybrid registration

Segmentation by registration has been proposed for the lung [4]. Since this method does not restrict atlas deformation, it can be used for segmenting pathological images but may also produce unrealistic results. A hierarchy of affine and non-rigid registration proposed by Rueckert et al [5] is used to calculate the transformation that warps the atlas 3D model to fit the subject's anatomy. We propose a hybrid registration process for training data set building. The method is optimised to match the characteristics of the prostate data and manual identification of a number of landmarks is required. Details of the stages and choices could be found in [6].

2.2 SDM registration

Using the results of hybrid registration, we perform principle component analysis on the positions of the control points to obtain a statistical model of the deformation - the statistical deformation model (SDM). Using a SDM in non-rigid registration for brain data was proposed by [3]. A number of advantages of SDM over a surface-based statistical shape model are suggested. SDMs can be constructed directly from images and do not require any prerequisite such as segmentation. SDMs allow the construction of an image atlas of the average anatomy as well as its variability across a population of subjects. Finally, SDMs take not only surface but the underlying anatomy into account by analyzing dense 3-D deformation fields.

We aim for detailed and accurate prostate segmentation for augmented reality guidance of robotic prostatectomy. However, since each mode of the SDM can affect the whole image space, the registration process is computationally expensive. One of the most time consuming parts of the non-rigid registration is gradient evaluation process due to the huge number of degrees of freedom. In the normal b-spline registration, a local region only is affected by motion of a node and this can be used for efficient gradient calculation [5]. Although the SDM reduces the degrees of freedom, the area of influence of each degree is extended, potentially to the whole image.

With a careful consideration of the complexity, SDM guided registration can prove to be a very efficient method. Taking into consideration previous works in the relevant fields, we propose, implement and evaluate two different methods to use SDM in the registration process - gradient projection method and SDM gradient with stochastic sub-sampling. In evaluating our implementations of automatic SDM registration, we consider both the algorithm efficiency and the accuracy improvement achieved.

2.2.1 Gradient projection We have adapted the original b-spline non-rigid registration process to incorporate the SDM 1. The gradient is calculated at the node point positions in the normal way. This gradient is then projected onto the SDM modes to provide the gradient in mode space. This should be almost as fast as the original registration process, since the projection is relatively a small computation compared to the gradient evaluation. The gradient projection should ensure that the result is only along the SDM mode directions and should encourage feasible transformations.

2.2.2 Gradient evaluation using stochastic sub-sampling within an interest region In order to estimate the gradient in SDM space, we must first know how many modes we should retain from the data set. A mathematical framework has been proposed to analyze the sources of PCA model error and minimizes over-modeling while keeping as much structural variance as possible to retain principal modes in [7]. This technique will be used to judge the completeness of the training dataset in our project.

Stochastic sub-sampling in non-rigid registration was first introduced by Klein et al [8]. In their approach a new and randomly selected subset of voxels is selected in each iteration of the optimisation process. We proposa that using

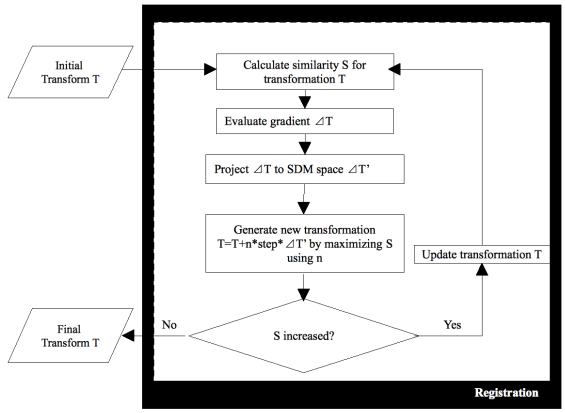


Fig. 1. Gradient projection registration process

the SDM could further improve the performance of the method while maintaining the speed benefit. The advantage of combining stochastic sub-sampling with SDM gradient evaluation is that we can use prior knowledge from the SDM to ensure sampling will be around the most significant nodes for a given mode. Another advantage is that the gradient is calculated in the SDM feature space rather than the space of b-spline control points positions.

Given an SDM with a mean vector M and modes $V = v_1, v_2, v_3, v_4 \ldots v_n$. For each mode $v_i = (var_1, var_2, var_3, \ldots, var_m)^T$, we use the following process to select the interest region. We order the control points from largest value to smallest value for the given mode until and find the smallest t such that $\sqrt{\sum_{i=1} 0^t var_i^2} \geq w$. This provides a set of control points of interest. All voxels controlled by the chosen control points are considered as the interest region. Since a cubic b-spline is used for calculation these voxels come from a 4x4 neighbourhood around the control points of interest. In our abdominal samples with w set to 0.95, half of the control points will be removed in the selection process.

The result is that all modes are normalized to get the gradient of modes in SDM space. Modes with a gradient which is less than a given fraction are set to be passive to further save computation time, with a possibility of $0.2 + 0.2 * n$, the modes in the passive pool may be set to active again, where n is the number of iterations for which the mode has been passive. Finally we optimise along the given gradient until the similarity improvement is negligible.

3 Results

Experiments comparing the two methods were performed on our lower-abdominal data. There are 22 samples. Analysis of the compactness of lower-abdominal data 4 shows the modes are not highly compact.

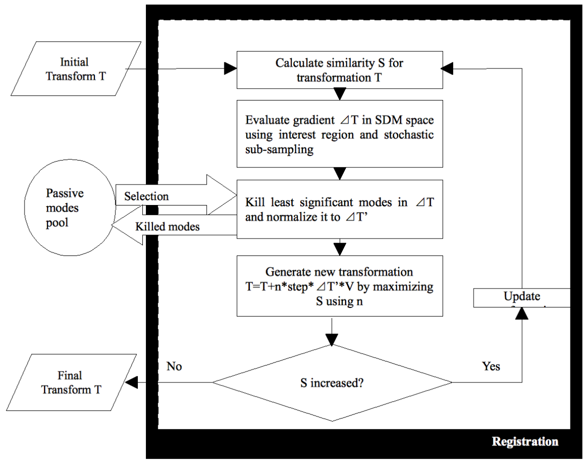


Fig. 2. Registration process of gradient evaluation using interest region and stochastic sub-sampling. Where V is the modes matrix in SDM

3.1 Complexity

As discussed in the previous sections, gradient projection adds a small extra computation to the gradient evaluation process. Gradient evaluation using SDMs feature space with interest region and stochastic sub-sampling has complexity is X times of normal B-spline registration, where $X = n * rdfactor * g(w) * F$ where n is number of modes we should retain from the data set, $rdfactor$ is the stochastic sub-sampling factor, $g(w)$ is the proportion of the interest region and F is the proportion of active modes. Experiments on the two methods were performed on our lower-abdominal data set.

	iterations	steps	time
Intensity nreg	4/5	4/4	67 mins
Gradient projection	2/5	2/4	42 mins
Stochastic sub-sampling	6/10	2/4	8 mins

Table 1. average terminate conditions experiment run on an Intel Due Core 1.8GHz Cpu 3gb Memory platform for lower abdominal data

Using 20 modes, the stochastic sub-sampling method works more than 8 times faster than the intensity only registration and gives improved accuracy.

3.2 Accuracy

The accuracy is measured in terms of the distance between the individual landmarks and reference landmarks which are selected manually to represent the

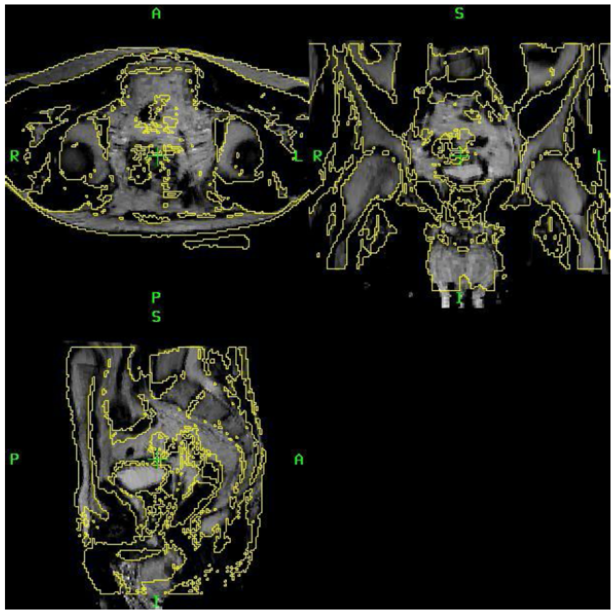


Fig. 3. An example stochastic subsampling registration result showing good anatomical alignment.

correspondence of the organs. There are 37 landmarks in all, distributed over the key position over the prostate, bladder and rectum. In figure 5 we can see some improvements on the overall alignment of the SDM based registration methods.

Figure 5 shows that all of the SDM guided registration methods have better performance in overall alignment of the individual images and reference images than intensity based non-rigid registration and are even better in some cases than the hybrid registrations result. This means that SDM have the power to integrate landmark information into the registration process or guide the registration using a more feasible space. In this way it is possible to achieve fully automatic registration and segmentation without the need for manual landmark identification. The resulting segmentation may be used as is or provide an initial starting point for manual segmentation.

4 Conclusions and Future Work

We firstly proposed a useful hierarchical segmentation-by-registration scheme for lower abdominal organs. Compared to manual segmentation, improvement is found at each stage and against intensity based registration. And with this scheme, models of patient could be built with very light labour work from MRI images [6].

Using the results from the above scheme, we built a statistical deformation model (SDM), and proposed two novel methods to use the SDM for guidance in

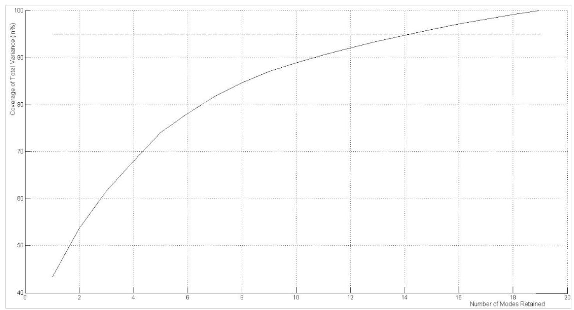


Fig. 4. Lower abdominal data sets SDMs compactness graph

the registration process. They are gradient projection and SDM gradient evaluation using an interest region and stochastic sub-sampling with passive mode pooling. Compared to manual segmentation using landmark correspondence, improvements in accuracy are found against intensity based non-rigid registration. The most encouraging result is that with our limited set of lower-abdominal data the SDM guided registration method with stochastic sub-sampling is 8 times faster then normal intensity based non-rigid registration and only takes 5-10 minutes. Our method also requires no manual intervention.

Due to the limitations of time and a relatively small training dataset, the potential benefits of the proposed method are not fully realised. The accuracy is not yet sufficent for image guidance. The aim for guidance of prostatectomy would be of the order of 3-5mm. In our case we obtain accuracy of 5-10mm. The first step for further work will include building an accurate expert segmentation database for the lower abdomen, refining the mesh model of the reference images, and building an SDM on a much larger training set. We will aim for a trainig set of 100-200 samples, but will use the method of Mei et al to establish when sufficiency has been achieved [7].

If the accuracy of the proposed method could be proven, it may then also possible to use non-linear techniques to build a more compact SDM to further optimise our method. Another option would be to use a GPU for the registration process in order to achieve real time performance. Our algorithm for registration is a suitable candidate for GPU implementation, since a repeated series of actions is used to evaluate position of every virtual pixel in the world coordinate of two corresponding images. This problem is readily parallelised for GPU implementation.

Additional research will focus on model instantiation from live intraoperative imaging such as ultrasound. We will also study the relationship between shape and motion. We have already shown that SDM guided registration can provide efficient and fairly accurate segmentation without manual intervention. The final goal of this research is to achieve a fully automatic; accurate and real time atlas registration for augmented reality guidance of surgery in the lower abdomen.

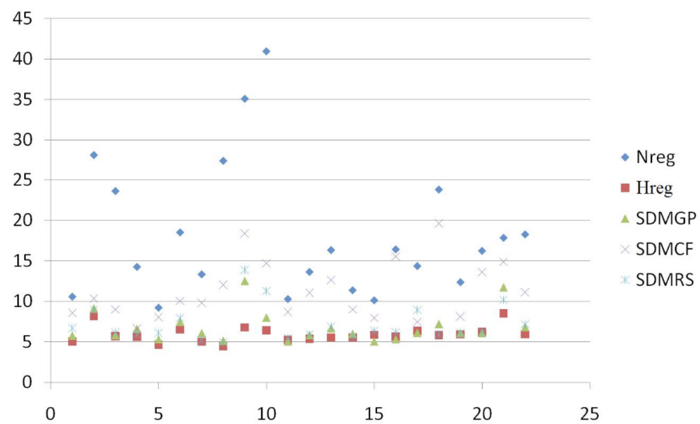


Fig. 5. Root Mean Square distance of landmark correspondence, last column is the average value. Nreg representing intensity based non-rigid registration, Hreg representing Hybrid registration, SDMGP representing Gradient projection registration, SDMCF representing cost function registration (not presented in this paper due to lack of space), SDMRS representing stochastic sub-sampling registration.

References

1. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Training models of shape from sets of examples. In: Proc. British Machine Vision Conference, Berlin, Springer (1992) 266–275
2. Park, H., Bland, P.H., Meyer, C.R.: Construction of an abdominal probabilistic atlas and its application in segmentation. IEEE Transactions on Medical Imaging **22** (2003) 483 – 492
3. Rueckert, D., Schnabel, A.F.J.: Automatic construction of 3-d statistical deformation models of the brain using nonrigid registration. IEEE Transactions on Medical Imaging **22**(8) (2003) 1014–1025
4. Sluimer, I., Prokop, M., vanGinneken, B.: Toward automated segmentation of the pathological lung in CT. IEEE Transactions On Medical Imaging **24** (2005) 1025 – 1038
5. Rueckert, D., Sonoda, L., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J.: Nonrigid registration using free-form deformations: Application to breast mr images. IEEE Transactions on Medical Imaging **28** (1999) 712–721
6. Shi, W., Mei, L., Darzi, A., Edwards, P.: Semi-automated segmentation of pelvic organs for AR guided robotic prostatectomy. In: Proceedings of AMI-ARCS 2008. (2008) 140–147
7. Mei, L., Figl, M., Darzi, A., Edwards, P.: Sample sufficiency and number of modes to retain in statistical shape modelling. In: Proceedings of MICCAI '08, LNCS. Volume 5241. (2008) 425–433
8. Klein, S., Staring, M., Pluim, J.P.W.: Evaluation of optimization methods for nonrigid medical image registration using mutual information and b-splines", 2007, vol. 16, pp. 2879-2890. IEEE Transactions on Image Processing **16** (2007) 2879–2890

Session 3 – Visualisation and Ultrasound

Intraoperative Ultrasound Probe Calibration in a Sterile Environment

Ramtin Shams[1], Raúl San José Estépar[2], Vaibhav Patil[2], and Kirby G. Vosburgh *[2,3]

[1] CECS, The Australian National University, Canberra
[2] Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA
[3] CIMIT, Boston, MA

Abstract. Strict sterility requirements impose restrictions on the type of equipment and its handling in an operating room (OR). This paper proposes three methods for intraoperative calibration of a tracked ultrasound probe under sterile conditions. We categorically call these methods *air calibration* to contrast them with the common phantom-based calibration methods that employ a coupling medium (e.g. a water-bath). The methods each consist of a preoperative and an intraoperative calibration stage. The preoperative stage is performed once in a non-sterile environment where any of the existing calibration methods can be used. The intraoperative stage is performed before each intervention in the OR. To minimize impact on the interventional work-flow, we required that the intraoperative calibration took less than 10 minutes, produced a robust result and was easy to perform.

1 Introduction

We have developed and demonstrated navigation systems for laparoscopic [1, 2] and endoscopic [3] transgastric interventions. The system has been validated *in vivo* on porcine models with a non-survival protocol [4] where sterility of the operating room (OR) is not a requirement. In moving from porcine to human subjects, sterility of the experimental equipment became a primary concern. This imposed limitations on the type of equipment and its handling in the sterile field (and during the disinfection process for endoscopy). As we have found, introduction of new material and practices into an OR is complex; it requires extensive iterative development of appropriate protocols and eventual certification. The process can postpone experiments for months. In this paper, we discuss one specific challenge that we encountered with the calibration of ultrasound and new methods that were devised around these practical limitations.

The navigation system consists of a laparoscopic ultrasound (LUS) or an endoscopic ultrasound (EUS) probe. An electromagnetic (EM) sensor (Ascension Technology) is attached close to the transducer on the probe where the relative position of the sensor w.r.t. the ultrasound plane can be maintained. In our

* Supported by NIH awards (2R42CA115112-02 and 1U41RR019703-01A2), CIMIT, and US Army (W81XWH-07-2-0011). The information does not necessarily reflect the position of the government, and no official endorsement should be inferred.

prior porcine experiments, we used Polyolefin tubes (shrink wrap) to secure the EM sensor. We then performed a standard *single-wall phantom* calibration [5] in a water-bath prior to the operation. The calibration procedure need not be repeated as long as the EM sensor is not removed.

In human subjects, however, there is a concern that biological sediments may not be properly removed by the disinfection or sterilization process, if the sensor is not detached from the instrument. This means that the EM sensor may not be configured permanently, but rather is attached after the components are sterilized separately and supplied in the sterile field of the OR. As such, there is a need for intraoperative calibration of the instrument in the sterile field.

An intraoperative calibration method has been proposed by Chen et al. in [6], who designed a *double-N* phantom that can be disassembled for sterilization and reassembled for the operation. Sub-millimeter accuracy is reported in conjunction with an optical tracking system.

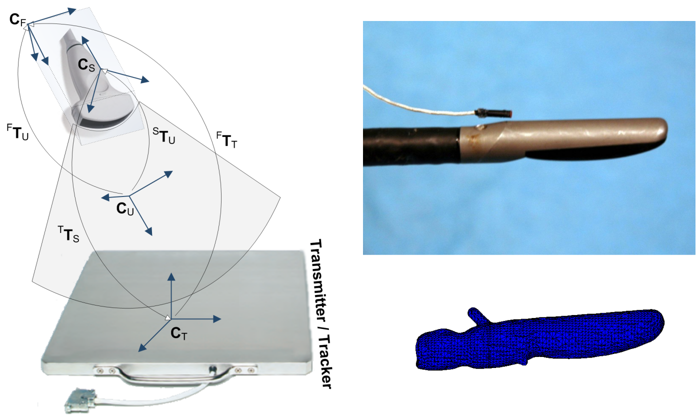
There are good arguments against ultrasound calibration in the OR with phantoms and liquids. Sterilization, phantom construction, and phantom assembly issues aside, accurate calibration is time-consuming and a delicate task. For example, correction for the speed of sound [7] may involve controlling the water temperature or creating water-glycerol or water-ethanol solutions. The more complicated an engineering solution, the less likely it is to be integrated into a clinical work-flow.

A phantom-less method for quality-control of calibration parameters is given by Boctor et al. [8]. The method can recover a sub-set of calibration parameters. In [9], Wein and Khamene propose a method that employs spatial consistency of two orthogonal freehand ultrasound sweeps of a region of interest to determine calibration parameters. The method is suitable for transcutaneous ultrasound probes and cannot be readily applied to the calibration of endoscopic or laparoscopic ultrasound.

2 Air Calibration

We propose three calibration methods (*Mold Calibration, Funnel Calibration,* and the *Closest Point Calibration*) that do not involve use of a phantom or a coupling medium (such as a water-bath) and can be performed easily by technicians in the OR. The third method has the added advantage that it uses only material already available in the OR and thus does not require an approval process. We categorically call these techniques *air calibration* methods.

These methods partition the process into preoperative and intraoperative calibration steps. Preoperatively, the ultrasound probe can be calibrated using any phantom-based method such as the single-wall phantom, Cambridge phantom, 3-wire phantom [5], or Hopkins phantom [10]. In a conventional calibration setup the following coordinate systems are defined, the tracker (world) coordinates $\mathbf{C}_{\mathrm{T}}$, the sensor coordinates $\mathbf{C}_{\mathrm{S}}$, and the ultrasound image coordinates $\mathbf{C}_{\mathrm{U}}$ as shown in Fig. 1(a). The conventional calibration problem determines ${}^{\mathrm{S}}\mathbf{T}_{\mathrm{U}}$ the transformation matrix from the ultrasound image's coordinate system to the sensor's coordinate system.



(a) (b)

Fig. 1. (a) Coordinate systems and transformations used in air calibration, (b) a laparoscopic probe and its 3D mesh model (needle partially visible in the biopsy channel).

For air calibration, we introduce an extra coordinate system ($\mathbf{C}_{\mathrm{F}}$) whose position remains fixed w.r.t. the ultrasound transducer. This coordinate system is attached to a physical (mold and funnel calibration) or virtual (the closest point calibration) object that maintains its relative position to the ultrasound plane.

We postulate that we can specify a stable coordinate system that remains fixed w.r.t. to the ultrasound coordinate system. The transformation matrix from the ultrasound plane to this coordinate system is denoted by ${}^{\mathrm{F}}\mathbf{T}_{\mathrm{U}}$. The purpose of the preoperative calibration is to determine this transformation. We first perform a conventional calibration and determine ${}^{\mathrm{S}}\mathbf{T}_{\mathrm{U}}$. Then, the position of $\mathbf{C}_{\mathrm{F}}$ is determined w.r.t. the tracker (${}^{\mathrm{T}}\mathbf{T}_{\mathrm{F}} = {}^{\mathrm{F}}\mathbf{T}_{\mathrm{T}}^{-1}$) and we can now determine ${}^{\mathrm{F}}\mathbf{T}_{\mathrm{U}}$ based on known relationships

$$ {}^{\mathrm{F}}\mathbf{T}_{\mathrm{U}} = {}^{\mathrm{F}}\mathbf{T}_{\mathrm{T}}\, {}^{\mathrm{T}}\mathbf{T}_{\mathrm{S}}\, {}^{\mathrm{S}}\mathbf{T}_{\mathrm{U}}. \tag{1} $$

During the intraoperative phase, the position of the EM sensor has changed (the sensor has been detached and reinstalled) and ${}^{\mathrm{S}}\mathbf{T}_{\mathrm{U}}$ from the preoperative stage is no longer valid and a new calibration is required. The intraoperative calibration stage consists of measuring the position of $\mathbf{C}_{\mathrm{F}}$ w.r.t. the tracker with the same method used in the preoperative step to determine ${}^{\mathrm{T}}\mathbf{T}_{\mathrm{F}}$ and then compute the calibration matrix ${}^{\mathrm{S}}\mathbf{T}_{\mathrm{U}}$ using known matrices

$$ {}^{\mathrm{S}}\mathbf{T}_{\mathrm{U}} = {}^{\mathrm{T}}\mathbf{T}_{\mathrm{S}}^{-1}\, {}^{\mathrm{F}}\mathbf{T}_{\mathrm{T}}^{-1}\, {}^{\mathrm{F}}\mathbf{T}_{\mathrm{U}}. \tag{2} $$

So far we have not been specific in setting the fixed coordinate system and its position and orientation in the coordinate frame of the tracking device. In the next sections, we discuss three air calibration methods to achieve this. The first two that require a device to be built are briefly discussed; the third method that uses point-to-surface registration will be explained in detail.

2.1 Mold Calibration

The tip of the ultrasound probe is cast to create a mold. We built a mold that fits the inflexible tip of the ultrasound probe. The asymmetries in the probe's shape ensure that the probe fits in the mold in a unique position. A second sensor is attached to the exterior of the mold. During the preoperative calibration stage, the ultrasound plane is calibrated relative to the mold sensor. The position of the mold sensor defines our fixed coordinate system, as the sensor does not have to be removed for sterilization (it does not directly touch the probe). During the intraoperative phase, a sensor is also attached to the probe and the probe is inserted in the mold. Probe calibration matrix is then computed using (2).

2.2 Funnel Calibration

This method is a variation of the mold calibration process, in which the mold is built with a sharp pointed end, like a funnel. Preoperatively, the ultrasound plane is calibrated against the tip of the funnel in a manner similar to calibration of a stylus. Since the position of the funnel tip does not change w.r.t. the ultrasound plane, the same process can be repeated intraoperatively to recover the calibration matrix. The funnel calibration requires a single sensor.

2.3 Closest Point Calibration

The previous methods require purpose-built appliances. Even when the appliances are built using approved material and are sterilizable, they must pass the approval process before they can be used. To further simplify the calibration process, we developed a third method that relies solely on material and devices already approved for use in the OR.

We use the same principle that the ultrasound probe must be calibrated w.r.t. a fixed point in space in relation to the ultrasound plane. We then create a model of the ultrasound probe by imaging the probe in a CT scanner and extracting a surface model by segmenting the image. Fig. 1(b) shows a laparoscopic probe and the surface model of its tip derived from a CT scan of the probe.. Three easy-to-identify landmarks (e.g. for laparoscopic probe: tip of the probe, entrance of the biopsy channel and lower exit of biopsy channel) are approximately identified in the model. The user is later instructed to touch these landmarks to initialize the registration process. We inserted a needle in the biopsy channel to give the model more spatial extent. We segmented the needle and the probe separately. The model with the needle inserted is used to guide the registration algorithm in the initial phase of the calibration process.

Preoperative Calibration: The preoperative calibration involves a calibration phantom, two EM sensors, and the probe to be calibrated. One of the sensors is mounted on the probe and the second one is used to scan the surface of the probe. In addition to the coordinate systems defined at the beginning of Section 2 we also have the coordinate system of the scanning sensor which we denote by $\mathbf{C}_{\mathrm{S}'}$. The fixed coordinate system $\mathbf{C}_{\mathrm{F}}$, in this setup is defined at an arbitrary (but fixed) position in the segmented CT volume.

We first determine the calibration matrix between the probe sensor and the ultrasound plane ${}^{\mathrm{S}}\mathbf{T}_{\mathrm{U}}$ using the single-wall phantom. We then scan the surface of the probe by moving the second senor slowly against the surface of the probe. The scanning sensor is attached to a stylus for easy handling.

The precise location of the coordinate system attached to the scanning sensor is not known. The offset between the location of the coordinate system and the tip of the sensor can be described by a translation. The position of the tip of the sensor $\mathbf{x}_t$ based on the sensor measurements $\mathbf{x}_{s'}$ is given by

$$\mathbf{x}_t = \mathbf{x}_{s'} + [v_x \ v_y \ v_z]^{\mathrm{T}}, \tag{3}$$

where the unknown translation vector $\mathbf{v} = [v_x \ v_y \ v_z]^{\mathrm{T}}$ is computed as part of our calibration/registration process.

The set of points measured on the surface of the probe $\{\mathbf{x}_t\}$ are related to corresponding surface points in the model $\{\mathbf{x}_m\}$ by a rigid transformation. The iterative closest point (ICP) algorithm can be used to register the two point sets as long as $\mathbf{v}$ is given. The adaptation of ICP to solve for $\mathbf{v}$ in addition to the rigid registration parameters is not trivial. One could install the sensor on a stylus and calibrate the stylus to retrieve $\mathbf{v}$. We did not find this option convenient nor particulary helpful for overall accuracy. The error is further compounded by the stylus calibration. Alternatively, we solve for $\mathbf{v}$ and the rigid registration together using a 9-parameter local optimization algorithm with a closest point cost function (a suitable optimization algorithm such as Powell, Simplex, or Gradient Descent variants can be used). The outcome of the closest point optimization is the transformation matrix from the tracker coordinates to the model (fixed) coordinates ${}^{\mathrm{F}}\mathbf{T}_{\mathrm{T}}$. We also measure the position of the sensor attached to the probe and can now compute ${}^{\mathrm{F}}\mathbf{T}_{\mathrm{U}}$ using (1). Note that the probe must remain stationary during the scanning process for this method to work. However, it is more convenient to have the flexibility to move the probe to gain access to the surface. The ability to move the probe has the added advantage that one does not have to worry about securing the probe in position and errors due to pressure against the tip of the probe which may cause small movements. To this end, we record the position of the scanning sensor and the probe sensor simultaneously, and compute the position of the surface points in the coordinate system of the probe sensor:

$$\mathbf{x}_s = {}^{\mathrm{T}}\mathbf{T}_{\mathrm{S}}^{-1}\,\mathbf{x}_t \tag{4}$$

Using $\{\mathbf{x}_s\}$ instead of $\{\mathbf{x}_t\}$ means that the outcome of the registration process is ${}^{\mathrm{F}}\mathbf{T}_{\mathrm{S}}$ and (1) can be simplified to

$${}^{\mathrm{F}}\mathbf{T}_{\mathrm{U}} = {}^{\mathrm{F}}\mathbf{T}_{\mathrm{S}}\ {}^{\mathrm{S}}\mathbf{T}_{\mathrm{U}}. \tag{5}$$

Intraoperative Calibration: The intraoperative calibration involves two EM sensors and the probe to be calibrated. The calibration process is similar to the preoperative calibration except that no phantom-based calibration is involved. The surface of the probe is scanned using a sensor and the resulting object measurements are registered against the model to determine ${}^{\mathrm{F}}\mathbf{T}_{\mathrm{S}}$, as before. Since ${}^{\mathrm{F}}\mathbf{T}_{\mathrm{U}}$ is known, the calibration matrix is computed using:

$$ {}^{\mathrm{S}}\mathbf{T}_{\mathrm{U}} = {}^{\mathrm{F}}\mathbf{T}_{\mathrm{S}}^{-1}\,{}^{\mathrm{F}}\mathbf{T}_{\mathrm{U}}. \tag{6} $$

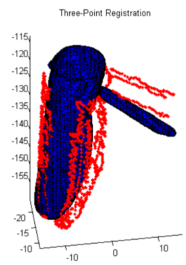
Object to Model Registration: To ensure convergence, a three-stage registration is performed for object to model registration. Each stage is designed to improve the registration accuracy and is initialized from the solution returned by the previous stage.

1. **3-point registration:** the user is requested to identify 3 previously selected landmarks in a pre-defined order by touching the corresponding points on the object by the sensor. An approximate rigid transformation from the object to the model is computed and used for initializing the registration optimization algorithm.
2. **9-parameter registration - initial:** a 9-parameter registration consisting of rigid object-to-model alignment parameters (6 parameters) and position of the sensor's tip in the coordinate system of the sensor (3 parameters) is performed. The results are used to initialize the next registration stage.
3. **9-parameter registration - final:** data points that belong to the needle are removed from both the object samples and the model and a constrained 9-parameter optimization is performed to determine the registration parameters more accurately.

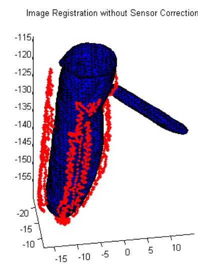
Fig. 2 illustrates incremental improvement in registration of a point of clouds measured from the surface of a laparoscopic probe (shown in red) to a model of the probe (shown in blue). Fig. 2(b) shows the result of a 6-parameter registration which does not include the calibration parameters of the sensor. This is shown for comparison with 9-parameter registrations only and is not part of our registration algorithm.

3 Results

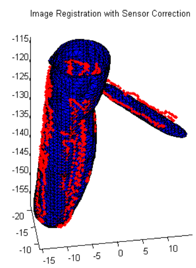
An LUS probe was calibrated preoperatively using the single-wall phantom. The probe was then registered to its 3D segmented model to compute the calibration matrix w.r.t. a fixed point in the model. For the intraoperative calibration we used a different sensor which was placed at a different location on the surface of the probe. The calibration was determined by the closet point calibration algorithm. Table 1 shows air calibration precision for 6 experiments. The first three rows show the mean registration error for each registration step. The registration error is reduced by each step. To validate the air calibration method, the intraoperative sensor was also independently calibrated using the single-wall phantom so that the air calibration results can be compared with the single-wall phantom. Single-wall calibration was also performed several times. The results are summarized in Table 2. The precision of the calibration methods was computed for a point in the center of the ultrasound image. This makes sense as the operators tend to keep objects of interests in the center of the field of view. We also report the preoperative calibration precision with the single-wall phantom for completeness.



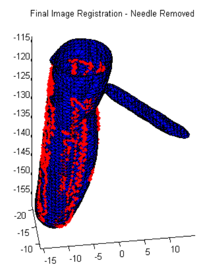
(a) 3-point registration



(b) 6-param. registration



(c) 9-param. registration



(d) 9-param. reg., needle removed

Fig. 2. Incremental improvement in the alignment of a scanned probe (red cloud) with the model (blue): (a) the registration is initialized with a 3-point rigid alignment first, (b) a 6-parameter registration is unable to register the point cloud to the surface of the model due to the distance of the origin of the sensor's coordinate system from its tip, (c) a 9-parameter optimization algorithm retrieves rigid registration parameters together with the sensor's calibration, (d) starting from the results of the previous run, a second 9-parameter optimization is performed with the needle points removed for improved alignment.

4 Discussion

This study demonstrates a fast, easy-to-use method for instrument calibration suitable for intraoperative use. It uses equipment and material already available in the OR. Our experiments were not directed toward reconstruction of 3D freehand ultrasound volumes but with approximate alignment of the B-mode ultrasound with a preoperative CT. The reformatted CT were shown side-by-side with the ultrasound stream to provide anatomical context for interpreting the ultrasound and improving the navigation of laparoscopic and endoscopic ultrasound [3]. For this application, the highest calibration accuracy was not the primary concern and we limited ourselves to qualitative assessment of the resulting calibration. It will be interesting to investigate the limits of the proposed methods for freehand ultrasound and to provide accuracy results in addition to precision in the future work. Single-wall phantom preoperative calibration is easy to perform but not the most accurate method. We expect the overall precision and accuracy to improve with a better preoperative calibration method.

Table 1. Registration error and the calibration precision for a number of experiments

Experiment	1	2	3	4	5	6
Mean reg. err. (step 1)	2.14 mm	1.60 mm	1.97 mm	3.20 mm	1.88 mm	2.81 mm
Mean reg. err. (step 2)	0.53 mm	0.48 mm	0.59 mm	0.60 mm	0.81 mm	0.64 mm
Mean reg. err. (step 3)	0.50 mm	0.48 mm	0.59 mm	0.54 mm	0.73 mm	0.63 mm
Air calibration error	2.07 mm	2.33 mm	0.37 mm	2.18 mm	2.27 mm	2.20 mm

Table 2. Precision of the single-wall phantom and the closest point calibration methods

Method	Err. Mean	Err. Std	Err. Min	Err. Max
Wall-phantom calib., pre-op. sensor	1.26 mm	0.44 mm	0.79 mm	1.85 mm
Wall-phantom calib., intra-op sensor	1.17 mm	0.40 mm	0.66 mm	1.71 mm
Air calib., intra-op sensor	1.90 mm	0.76 mm	0.37 mm	2.33 mm

References

1. Ellsmere, J., Stoll, J., Rattner, D.W., Brooks, D., Kane, R., Wells III, W.M., Kikinis, R., Vosburgh, K.G.: A navigation system for augmenting laparoscopic ultrasound. In: Medical Image Computing and Computer Assisted Intervention (MICCAI). (November 2003) 184–191
2. Ellsmere, J., Stoll, J., Wells III, W.M., Kikinis, R., Vosburgh, K.G., Kane, R., Brooks, D., Rattner, D.: A new visualization technique for laparoscopic ultrasound. Surgery **136**(1) (July 2004) 84–92
3. Vosburgh, K.G., Stylopoulos, N., San José Estépar, R., Ellis, R.E., Samset, E., Thompson, C.C.: EUS with CT improves efficiency and structure identification over conventional EUS. Gastrointestinal Endoscopy **65**(6) (2007) 866–870
4. San José Estépar, R., Stylopoulos, N., Ellis, R.E., Samset, E., Westin, C.F., Thompson, C.C., Vosburgh, K.G.: Towards scarless surgery: An endoscopic ultrasound navigation system for transgastric access procedures. Computer Aided Surgery **12**(6) (2007) 311–324
5. Prager, R., Rohling, R., Gee, A., Berman, L.: Rapid calibration for 3-D free-hand ultrasound. Ultrasound in Med. & Biol. **24**(6) (July 1998) 855–869
6. Chen, T.K., Thurston, A.D., Ellis, R.E., Abolmaesumi, P.: A real-time freehand ultrasound calibration system with automatic accuracy feedback and control. Ultrasound in Med. & Biol. **35**(1) (January 2009) 79–93
7. Mercier, L., T., Langø, Lindseth, F., Collins, L.D.: A review of calibration techniques for freehand 3D ultrasound systems. Ultrasound in Med. & Biol. **31**(4) (April 2005) 449–471
8. Boctor, E.M., Iordachita, I., Fichtinger, G., Hager, G.D.: Ultrasound self-calibration. In: Proc. SPIE Medical Imaging: Visualization, Image-Guided Procedures, and Display. (February 2006)
9. Wein, W., Khamene, A.: Image-based method for in-vivo freehand ultrasound calibration. In: Proc. SPIE Medical Imaging: Image Processing. (February 2008)
10. Boctor, E.M., Jain, A., Choti, M.A., Taylor, R.H., Fichtinger, G.: Rapid calibration method for registration and 3D tracking of ultrasound images using spatial localizer. In: Proc. SPIE Medical Imaging: Visualization, Image-Guided Procedures, and Display. (February 2003) 521–532

Realtime Integral Videography Auto-stereoscopic Surgery Navigation System using Intra-operative 3D Ultrasound: System Design and In-vivo Feasibility Study

Nicholas Herlambang[1], Hiromasa Yamashita[1], Hongen Liao[2], Ken Masamune[1], and Takeyoshi Dohi[1]

[1] Graduate School of Information Science and Technology,
[2] Graduate School of Engineering, The University of Tokyo, Japan
nicholas@atre.t.u-tokyo.ac.jp *

Abstract. 3D ultrasound is non-invasive and fast imaging method suitable to guide surgeries with large intra-operative organ movement and deformation. But it is usually displayed on 2D screen so depth information is difficult to percept. In this paper, we present real-time Integral Videography (IV) auto-stereoscopic surgery navigation system using intra-operative 3D ultrasound. The system visualized 3D ultrasound data as high quality auto-stereoscopic images in real-time. We improved IV image quality by adding Phong shading into original composite algorithm, implemented on GPU for real-time calculation. Comparison with original method showed that the use of Phong shading for IV improved depth perception and reality of IV images, with performance decrease of less than 50%. As clinical feasibility study, we conducted an in-vivo porcine experiment simulating mitral valve surgery on beating heart, guided with real-time auto-stereoscopic image of intra-operative 3D ultrasound. The experiment showed that our system is fast enough to follow heart beat. Surgical tool was visible clearly and succesfullly driven towards surgery target. We also received qualitative evaluation from expert cardiologist that time lag was within tolerable range.

1 Introduction

3D Ultrasound is the fastest imaging technology capable of acquiring 3D images of human body in real-time. This capability, along with the fact that ultrasound is non-invasive, has brought 3D ultrasound to be used for diagnostic and surgeries. Unfortunately, 3D images acquired are usually displayed as a 3D view on 2D screen. Displaying 3D view on 2D screen causes lack of depth perception. Surgeons may have to rotate the viewing angle of the 3D view to percept depth information. While it may be enough for diagnostics where there are plenty of time for decission, it is not practical for surgeries, especially for surgeries with

*

large organ movement and deformation where decision time is minimal, such as beating heart surgery. A more intuitive vizualizion way such as stereoscopic display is required.

We have developed an auto-stereoscopic visualization method called Integral Videography(IV)[1, 2]. The principle of IV to display a stereoscopic image is by projecting lights from high-resolution LCD screen onto 3D space in front of the screen by using micro convex lens array. Compared to other stereoscopic visualization method such as binocular stereoscopic, IV has many advantages, such that it doesn't need glasses or other viewing devices, it is visible by many people at the same time, and it is spatially accurate. The fact that it is spatially accurate makes it possible to use it in applications such as IV overlay, augmenting IV stereoscopic image with patient's body.

However, in order to bring IV into operating room, there are still many issues to solve. First, IV rendering, the calculation to create an IV stereoscopic image from 3D data is heavy computationally. Second, current IV rendering algorithm lack image quality. Third, there is a need for user interface for surgery navigation purpose. Herlambang et al. have developed a fast IV rendering algorithm using GPU implementation[3], and a composite IV rendering method for high image quality rendering[4]. Implementation of IV rendering algorithms on GPU has made it possible to display IV images in real-time. And composite rendering for IV has improved image quality significantly, though lack of shading algorithms is still room for improvements.

In this paper, we developed a real-time IV surgery navigation system using 3D ultrasound. In details, we added shading algorithm into composite IV rendering algorithm, develop the user interface, and finally performed an in-vivo feasibility study.

2 Method

2.1 System Configuration

Fig 1 shows system configuration of the real-time IV visualization system. It consists of an ultrasound device(α10, ALOKA), a workstation (Quad Core 2.4 GHz, 4GB), an LCD screen for GUI display, and an IV stereoscopic display. The ultrasound probe used is 3D convex probe that can acqurire 3D image up to 10 volumes/s, in trade-off with image resolution and quality. The 3D ultrasound data is sent to workstation in real-time through LAN and LVDS connection. Acquired 3D ultrasound data is then processed on the workstation, with help of GPU computing server, and the resulted IV image is displayed on the IV display. As the IV display we used a high resolution LCD screen (6.4 inch XGA). We used convex lens array with lens diameter of 1 mm arranged in hexagonal configuration.

2.2 IV Rendering Algorithm

IV rendering algorithm to create IV image from 3D data is based on ray-tracing algorithm. The image value of each pixel on LCD screen is calculated by ray

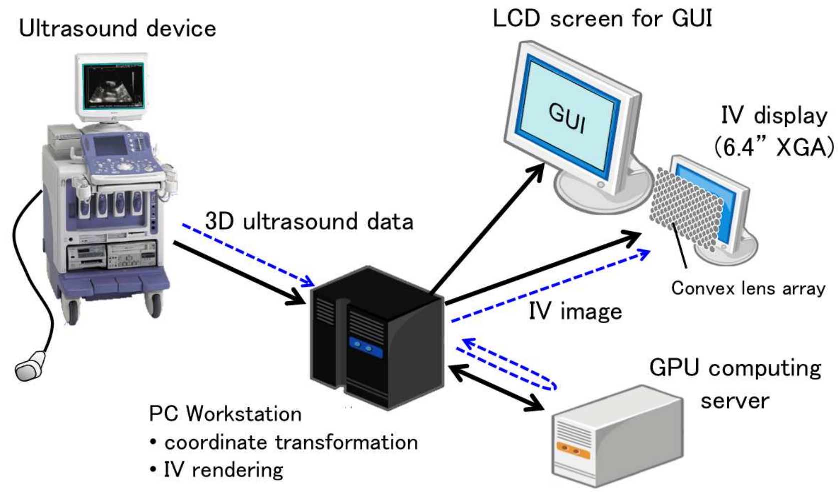


Fig. 1. System configuration.

tracing along the light ray connecting the pixel and center of corrensponding convex lens. We have developed composite rendering algorithm for IV, which use color and alpha transfer function to each voxel value, and use those values as color and transparency of each voxel when doing ray tracing. To improve previous algorithm, we implemented Phong shading algorithm[5]. Previous algorithm without shading only calculated ambient component of Phong shading, while the proposed algorithm also calculated diffuse and specular components (Fig 2). In Phong shading calculation, there is a need to also calculate gradient field of image data. Gradient field, also implemented on GPU and pre-calculated on each image update, is then used to calculate diffuse and specular components.

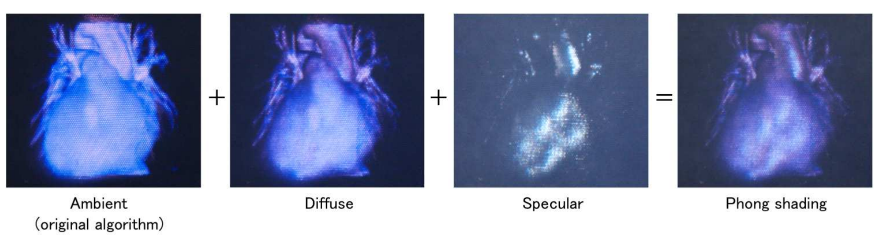
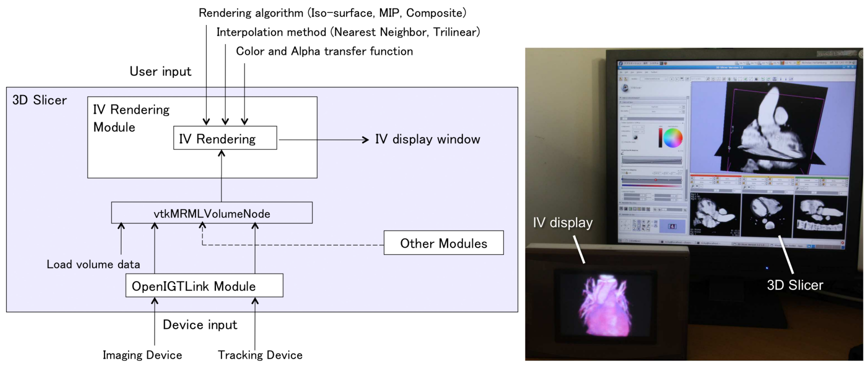


Fig. 2. Phong shading for IV stereoscopic image. In addition to ambient component of original algorithm, diffuse and specular component are also calculated.



(a) (b)

Fig. 3. GUI of IV visualization system as a module of 3D Slicer: extensibility and functionality. (a) procesing pipeline (b) IV rendering module in 3D Slicer

The ratio between weight coefficients of each component defines the nature of reflection. We expect the proposed IV rendering result to have superior depth perception compared to original algorithm, in trade-off with calculation time.

2.3 GUI

We developed IV visualization system user interface as an extension module of open source image processing software (3D Slicer[6]). The advantage of this approach is that it is easy to combine it with existing features and modules of 3D Slicer, and therefore it is easy to build various surgery navigation applications. The IV rendering module follows data processing pipeline of 3D Slicer by using MRML data node (Fig 3). IV rendering module takes vtkMRMLScalarVolumeNode as input, so any 3D image will be dealt with the same way. Any volume node that is inputed into IV rendering module will be observed for update so that any change to the volume node will trigger an IV rendering cycle. In visualization of 3D ultrasound data, IV rendering module is used in combination with OpenIGTLink module[7]. OpenIGTLink module receive 3D data in real-time from other application (in this case, 3D ultrasound data acquisition software), and keep it in 3D Slicer memory as a vtkMRMLScalarVolumeNode, such that when inputed into 3D Slicer it will result in real-time visualization of IV stereoscopic image. Other features such as thresholding, ROI clipping were also implemented. And also, there is an option to enable automatic rendering performance control. If enabled, step size in ray tracing is automatically adjusted to match desired rendering frame rate.

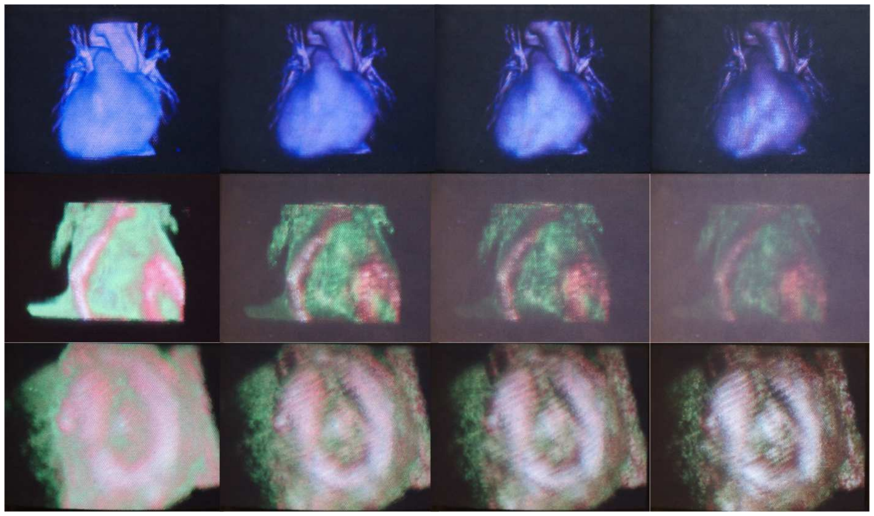


Fig. 4. IV stereoscopic visualization of various datasets using different shading parameters. top to bottom: CT data of human heart, MRI data of human heart, ultrasound data of porcine heart (mitral valve only). Left to right: IV rendering result with various shading parameter, Ambient:Diffuse:Specular = 1:0:0(original algorithm), 0.16:0.84:0, 0.08:0.75:0.17, 0.08:0.5:0.42.

2.4 Implementation Issues

Calculations of 3D ultrasound data transformation and IV rendering were implemented on GPU calculations using CUDA programming platform[8]. The workstation has an internal GPU for calculation and visualization (Quadro FX 5800, NVIDIA), and an external GPU computing server for calculation (Tesla D870, NVIDIA). To avoid resource usage timing conflict, 3D ultrasound data transformation and IV rendering were calculated on separate GPU. 3D ultrasound data transformation was performed on external GPU computing server, while IV rendering was performed on internal GPU.

3 Results

3.1 IV Image Quality Comparison

We compared IV image quality of the new IV rendering algorithm (composite rendering with Phong shading) with the original algorithm (composite rendering without shading). Targetting applications for heart surgery, We tried visualization on various datasets: CT, MRI, and ultrasound datasets. Visualization was compared between various shading parameters (Fig 4). In Fig 4, we put more weight on specular and diffuse components gradually from left to right. The CT

Table 1. Average calculation time of IV volume rendering

Data size	Composite (ms)	Composite with shading (ms)	% slower
64^3	23	23	0%
128^3	36	39	8%
256^3	59	77	31%
512^3	164	245	49%

screen size: 1024×768 pixels (n=100)

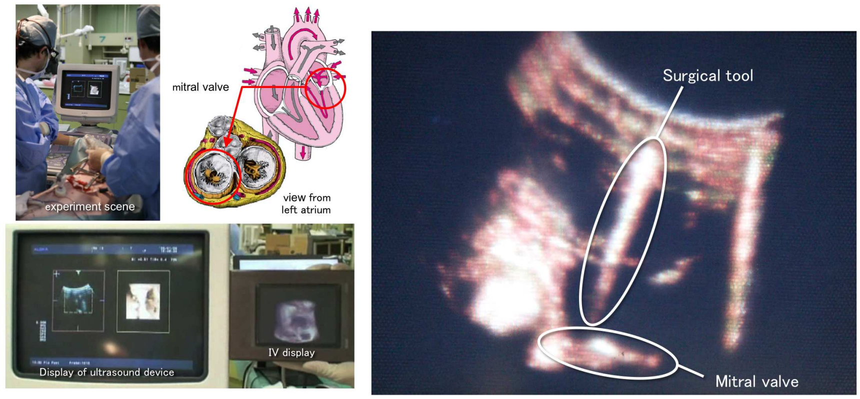
dataset was acquired using single slice CT with ECG gating. CT data size is 512×512 pixels $\times 192$ slices. MRI dataset was acquired using 0.2T MRI with heart pulse gating. Acquisition time was around 40 minutes, and since we did not perform respiratory gating, the MRI data was rather noisy due to motion artifacts. MRI data size is 256×256 pixels $\times 19$ slices.Ultrasound dataset was acquired with mechanical 3D convex probe, with data size of 304×248 pixels $\times 44$ slices and data acquisition rate of 3 volumes/s. In case of low noise CT-dataset, shading is smooth, and the more specular component put on, the better the depth perception. In case of rather noisy datasets of MRI and ultrasound data, too much weighing on specular component may result in decrease on overall image brightness and contrast.

3.2 Rendering Time Evaluation

We evaluated IV rendering time of the new IV rendering algorithm (composite with shading) compared with the original algorithm (composite) for various data sizes (Table 1). For small data size, the new algorithm perform as fast as the original algorithm, but for big data size the new algorithm is up to 49% slower than the original algorithm.

3.3 In-vivo Experiment

To demonstrate the usefulness of IV surgery navigation system, we performed an in-vivo porcine(male, 47.5 kg) experiment simulating a mitral valve surgery on beating heart navigated by IV images of 3D ultrasound (Fig 5(a)). The surgery was conducted by expert cardiologist. The most important requirement to use the IV system in heart surgery is that it should be fast enough to follow heart beat movement. Therefore, in this experiment, firstly we tested IV system to visualize mitral valve movement in real-time. 3D ultrasound data acquisition was performed for several combinations of resolution and data acquisition rate, up to 8 volumes/s. Acquired 3D ultrasound data is in polar coordinate system and then transformed into rectangular coordinate system before being used for IV rendering. For all cases, 3D ultrasound data is transformed into $256 \times 256 \times 256$ voxels data, which is rendered at around 13 fps, therefore IV rendering frame rate was still faster than data acquisition rate. So there was no frame skipping



(a) (b)

Fig. 5. In vivo porcine experiment: (a) Real-time IV stereoscopic image from intra-operative 3D ultrasound as image guidance. (b) simulating mitral valve surgery on beating heart, surgical tool was driven towards mitral valve under IV stereoscopic image guidance.

and time lag was less than 1 frame. Then, we guided a surgical tool towards mitral valve under 3D ultrasound guidance displayed as stereoscopic images on IV display (Fig 5(b)). For this first trial, we did not put any efforts on manipulator coating or by using manipulator with special material suitable for ultrasound data acquisition. With optimal rendering parameter setting, manipulator was seen clearly along with the beating heart, and it is easy to guide the manipulator towards mitral valve, and perform some surgery manipulations. According to the cardiologist conducting the surgery, time lag was within tolerable range.

4 Discussions

The use of Phong shading for IV rendering improved overall depth perception and image quality. However, based on imaging modality used, it is required to use modality specific shading parameters. For rather noisy datasets, smoothing on image gradient may improve shading quality. Decrease in performance was more significant for large dataset. The reason is that overall rendering time includes data transfer time between GPU and CPU that scales up with the number of voxels processed. In-vivo experiment showed that our system is feasible to guide open-chest beating heart surgery with 3D ultrasound. Our system is comparable with similar systems[9, 10] in terms of using real-time intra-operative images for guiding intra-cardiac beating heart surgery. Linte used 2D ultrasound augmented with pre-operative 3D CT image, while Li used 3 slices of real-time MRI images. Our system, using auto-stereoscopic 3D ultrasound images, should be the most intuitive among the three, and surgeries with fast moving target, such as mitral valve surgery should benefit the most from our system.

5 Conclusion

We have built a real-time autostereoscopic surgery navigation system with intra-operative data acquisition using 3D ultrasound. Implementation of Phong shading algorithm for IV has improve image quality and depth perception significantly. Performance decrease due to implementation of shading algorithm is less than 50%. It is still considerably fast enough to visualize 512^3 datasets in real-time. We also built GUI for IV visualization in modular design, so that it is straight-forward to combine it with various surgery navigation applications built on 3D Slicer. Clinical feasibility study on porcine phantom simulating open chest surgery showed that our system provide enough speed and image quality to guide beating heart surgery. In the future, we plan to use our system for wide range of applications, especially surgeries with large organ movements and deformations such as beating heart surgery and fetus surgery.

Acknowledgements

This work is partially supported by Research Fellowships for Young Scientists and the Grant-in-Aid for Scientific Research of the Japan Society for the Promotion of Science (JSPS).

References

1. Hongen Liao, e.a.: Intra-operative real-time 3d information display system based on integral videography. In: MICCAI 2001, LNCS 2208. (2001) 392–400
2. Hongen Liao, Nobuhiko Hata, T.D.: Image-guidance for cardiac surgery using dynamic autostereoscopic display system. In: IEEE ISBI 2004. (2004) 265–268
3. Nicholas Herlambang, e.a.: Interactive autostereoscopic medical image visualization system using gpu-accelerated integral videography direct volume rendering. In: CARS 2008. (2008) S110–S111
4. Nicholas Herlambang, e.a.: Real-time autostereoscopic visualization of registration-generated 4d mr images of beating heart. In: MIAR 2008, LNCS 5128. (2008) 349–358
5. Phong, B.T.: Illumination for computer generated pictures. Communications of the ACM **18**(6) (1975) 311–317
6. BWH: 3d slicer (Mar 2009) http://www.slicer.org.
7. NAMIC: Openigtlink (Mar 2009) http://www.na-mic.org/Wiki/index.php/OpenIGTLink.
8. NVIDIA: Cuda zone home (Mar 2009) http://www.nvidia.com/object/cuda_home.html.
9. Cristian A. Linte, e.a.: Virtual reality-enhanced ultrasound guidance for atrial ablation: in vitro epicardial study. In: MICCAI 2008, LNCS 5242. (2008) 644–651
10. Ming Li, Dumitru Mazilu, K.A.H.: Robotic system for transapical aortic valve replacement with mri guidance. In: MICCAI 2008, LNCS 5242. (2008) 476–484

A Movable Tomographic Display for 3D Medical Images

Gaurav Shukla[1], Bo Wang[1], John Galeotti[2], Roberta Klatzky[3], Bing Wu[3], Bert Unger[1,2], Damion Shelton[2], Brian Chapman[1], and George Stetten[1,2]

[1] Department of Bioengineering, University of Pittsburgh, Pittsburgh, PA, 15261
[2] Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, 15213
[3] Department of Psychology, Carnegie Mellon University, Pittsburgh, PA, 15213 *

Abstract. We describe here the first working prototype of a novel display for viewing 3D medical images. The position and orientation of a freely movable touch-screen display are optically tracked and used to continuously determine which slice to display within a 3D data set. The slice is registered "in situ" relative to a fixed coordinate system, through which the display is moved. We have coined the term "grab-a-slice" for the new display, to connote the intuitive nature of the interaction it provides with volumetric data, potentially more so than that provided by traditional fixed displays. With grab-a-slice, the user experiences the illusion of slicing through an invisible patient. The touch-screen allows the user to directly identify the location of any point of interest within the 3D image data. Grab-a-slice has a number of possible clinical and scientific applications. In particular, we are exploring its utility for improved vascular tracing to identify pulmonary embolus in contrast-enhanced computed tomography (CT). In addition, we are planning psychophysical studies of how users explore and navigate through medical image data with this new display. We are also developing methods of graphical augmentation for grab-a-slice using stereo display, to improve the ability of users to understand the raw content of a tomographic slice in the context of the surrounding 3D anatomy and to improve their ability to navigate through a 3D dataset. Finally, we are exploring the use of grab-a-slice to supervise semi-automated image analysis routines.

1 Introduction

Three-dimensional medical images, such as those acquired by magnetic resonance (MR), computed tomography (CT), or other volumetric imaging modalities have been a great boon to physicians seeking to learn non-invasively about a patient's condition. In clinical practice, 3D images are generally still viewed one slice at a time on a 2D display, using a mouse or keyboard to sequence through a stack of such slices. The slices are usually oriented orthogonal to the cardinal axes (sagittal, coronal, axial) representing samplings from the original 3D data, or

* We acknowledge NIH grants 1R01HL087119, RO1HL089750, and R21EB007721, which supported this work.

the slices may be arbitrarily oriented, computed by interpolating values from the original 3D data. In the current clinical reading room, the display screen remains motionless before the observer as the slice is moved through the patient. While this method of display has been sufficient for many applications, it creates a disconnect between the 3D anatomic relationships in the patient and the stationary sequence of 2D images, especially when the slices represented are not parallel, i.e., the view path is curved. Planning a procedure (e.g., image-guided needle biopsy) in the 3D coordinates of the patient may be less than intuitive when the images are displayed on such an immobile 2D screen.

To address this problem, we have constructed a special type of medical image display that is free to move about in a 3D space representing the coordinate system of the data. At any given time, the display shows a slice that corresponds to the current location and orientation of the display. We hypothesize that manipulating such a display through what amounts to an "invisible patient" will preserve the perception of 3D anatomic relationships in a way not possible with current immobile displays. We use the term "grab-a-slice" to describe this new type of display. We report here the first working prototype.

2 Related Work

Various methods have been developed for rendering 3D data onto a stationary 2D display, with or without special hardware for stereovision (e.g., Levoy's classic paper [1]), but these are not widely used by clinicians. Navigation through a 3D environment has also inspired several approaches. Ware and Osborne [2] provide a user interface for exploring virtual graphics environments they call "scene-in-hand," a virtual camera control that changes the perspective of 3D environment in response to the manipulation of a tracked tool. Hinckley et al. use passive interface props, [3] tracked objects that are simple and hand-held, to generate tomographic slices of medical image data. The cubic mouse, [4] developed for specification of 3D coordinates in graphics applications, consists of a tracked in-hand device coupled with rods and buttons to specify motion of virtual objects along various axes. Other approaches, including the SpaceBall line of products (3DConnexion, Silicon Valley, CA) implement a non-tracked 3D navigation tool. All of the above still used a stationary display, as opposed to the movable display described here.

Tracked movable boom-mounted displays [5][6] that are counterbalanced so that the operator can manipulate them by hand have been used as immersive displays into 3D virtual environments. However, to our knowledge, they have not been used for tomographic slicing of a volume.

Augmented reality systems in which head-mounted displays are coupled with algorithms for 3D perspective rendering have been studied extensively. [7][8]This approach has also been applied to viewing tomographic slices. [9][10]

The grab-a-slice display has evolved out of an effort in our laboratory and elsewhere to develop image guidance systems that merge ultrasound (US) images with a direct view of the patient. [11][12] Our device, called the Sonic Flashlight (SF), consists of a small display and a half-silvered mirror mounted on a conventional US transducer. Looking through the mirror, the operator sees the reflection of the real-time US image floating *in situ* within the patient, precisely where the scan is currently being obtained. The SF merges the US image, US probe, operator's hands, surgical instrument, and patient into the same field of view, enabling perceptually guided action. Similar approaches have been taken by others for displaying slices of CT and MR data *in situ*. [13–15]

We have conducted extensive research into the underlying psychophysical properties of *in situ* image guidance. This research has demonstrated advantages as compared to conventional displays in the accuracy of perceived target depth, immunity to errors due to surface deformation, and the interpretation of shape and pose of 3D targets. [16–20] We use the term "tomographic aperture" to denote the manner in which 3D data is sampled by slicing, analogous to that of a conventional aperture through which the world is sampled by projection. We have found that the *in situ* image display provides a perceptual link between in-plane and through-plane distance as well as a spatial buffer for memory to combine sequential 2D information into a 3D context.

This avenue of research led us to conceive of the tracked grab-a-slice display, which produces an *in situ* image with larger size and greater clarity than possible with the SF. As opposed to the SF's virtual image, which is limited by the intervening mirror, grab-a-slice represents an easily manipulable tomographic aperture based on a real image that can be touched with a finger to identify individual points in 3D space.

3 Methods

We constructed a wooden frame for a 15" touch-screen display (Microtouch M150, 3M, Inc.) that allows the display to be manipulated in 4 degrees of freedom (DOF) with respect to a stationary tabletop. A user may move the display with a single hand using one of two handles (see Figure 1). The apparatus is free to translate across the tabletop in two directions (A and B), and to rotate about the "yaw" axis (C), facilitated by Teflon pads under the platform, which provide low dynamic friction and allow for comfortable one-handed manipulation. A hinged mount with ball-bearing tracks permits the display to rotate about the "pitch" axis (D). Sufficient static friction in the hinge and between the platform and the tabletop, along with proper balance of the hinged assembly, guarantees that the screen remains immobile when released. Rigid bodies have a total of 6 DOF, but we deemed it unnecessary to physically implement the remaining 2 DOF: translation in height normal to the tabletop and rotation in the "roll" direction within the image plane. Both can be readily implemented in software, and a footswitch can control, in effect, raising and lowering the invisible patient or rolling the patient around the long axis from prone to supine. A seventh DOF, isotropic scale, can also be manipulated in software, effectively magnifying or shrinking the entire patient.

Fig. 1. The current grab-a-slice display can be manipulated in four degrees of freedom (see text).

We mounted 10 infrared light-emitting diode (IRED) markers on the grab-a-slice display for detection by an optical tracking system (Optotrak Certus, Northern Digital Inc.). The Optotrak can localize each IRED marker with an accuracy of approximately 0.1mm and a sampling frequency of at least 100 Hz, using a rigid array of three cameras fixed with respect to the tabletop. All the markers are mounted on the portion of the grab-a-slice apparatus that is rigidly attached to the display itself. Thus the markers can be treated as a rigid body by the Optotrak software to compute orientation and position for the display as a whole relative to the camera array.

For our system to function, each point in the 3D image must correspond uniquely to a point in physical space. This is achieved by placing the image data at a fixed location relative to the camera origin. We adapted software originally designed in our laboratory for displaying simulated and pre-acquired 3D data with the SF. [21] The software takes measurements from the optical cameras and performs the image slicing and rendering, all in real time.

Calibration of the grab-a-slice consists of the following procedure. An optically tracked stylus is used to find the 4 corners of the touch-screen display in camera coordinates, and the scale of the displayed data is adjusted accordingly so that the data is displayed as life-size (the scale may also be changed intentionally). The grab-a-slice display is assigned to be at the mid-axial slice across the thorax when placed at the center of the table with zero pitch and zero yaw. The long-axis of the patient is assigned to the "range" or z axis of the camera coordinate system.

As the apparatus is moved to other locations and orientations, the Optotrak software continually reports the locations of the 4 corners of the display based on their relationships to the IRED markers. To display the appropriate slice

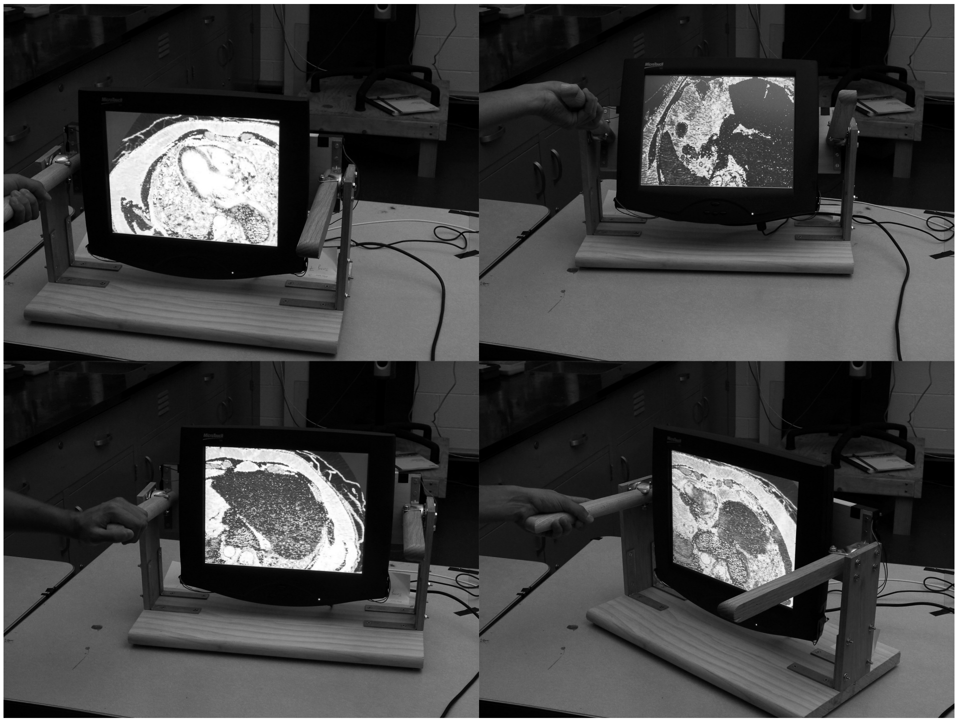


Fig. 2. The grab-a-slice display showing CT images of the lung. Various locations and orientations of the display result in corresponding slices through the invisible patient being displayed. Degrees of freedom consist of 2 translations and 2 rotations.

from the 3D data on the display, the locations of the 4 corners of the display are used to extract the appropriate slice from the 3D data set, by means of a 3D texture mapping board (GeForce 8800, NVIDIA, Inc.). The method of 3D texture mapping interpolates voxels from a 3D data set onto polygons in arbitrary planes for 2D display, in this case a single rectangle occupying the surface of the touch-screen. [22]

A 3D dataset consisting of a (de-identified) CT scan of a human thorax was used for our initial demonstration. The results are shown in Figure 2. As can be seen, the displayed image content responds to movement of the apparatus through the coordinate system of the CT image data.

Since the display is also a touch-screen, the operator can easily and unambiguously identify locations in the image coordinate system. This feature serves as an intuitive 3D mouse. A separate immobile display (not shown) operates as a control panel to activate the system, select the particular 3D data set to load, as well as adjust brightness, contrast, and the additional transform parameters discussed above.

4 Discussion

Grab-a-slice represents a potentially useful tool in medicine. We envision the device to have at least three areas for clinical application.

One area is preoperative planning. Surgeons often examine medical images before performing a procedure. For example, a surgeon attempting to extract a bullet from a patient's abdomen may look at an abdominal CT scan to identify the bullet's location, possible obstacles, and to determine an optimal path of approach. At present, the surgeon must cognitively relate this information, displayed as 2D slices on a stationary screen, to the subsequent 3D interaction with the patient in the operating room. Being able to examine the CT data with grab-a-slice may provide the surgeon with a more intuitive sense of 3D anatomical relationships, analogous to what is termed in the military as, "situational awareness." The surgeon can preoperatively plan and record a surgical path and demonstrate it to the surgical team. The screen may be located on a table directly adjoining the patient, with the "invisible patient" oriented parallel to the real one, thus making corresponding orientations and distances directly comparable.

A second clinical application involves using grab-a-slice as a training tool. Medical students, nurses, residents, and other health professionals often have difficulty learning to interpret 3D medical images; the orientation of the patient and the relationship between slices is not always readily apparent. Medical students in the anatomy laboratory could dissect a liver in the cadaver while examining a grab-a-slice rendering of a CT scan of the same liver on an adjacent table, gaining expertise with medical images at an early stage in their training.

A third clinical realm for grab-a-slice is diagnostic radiology. The particular application we are studying is the diagnosis of pulmonary embolism (PE), an acute, life-threatening, and treatable condition with over 650,000 incident cases in the USA annually. PE is diagnosed by a combination of clinical symptoms, laboratory results, and medical imaging to determine the presence, location, and size of potential emboli. Radiological evaluation of CT data for evidence of PE involves tracing branches of the pulmonary vessels containing suspected emboli back to the heart, to determine whether the vessel in question is an artery or a vein (pathologic emboli are always in the arteries). Tracing these vessels along their course is a non-trivial task using a stack of 2D slices, and we are evaluating whether it may be made easier and more reliable with grab-a-slice.

In addition to offering potentially useful clinical applications, grab-a-slice represents a new platform for at least two areas of fundamental psychophysical research: 3D visualization and 3D navigation. The processes that humans use to build up mental representations of 3D structures are of great interest to the psychophysical community. Grab-a-slice could be used to study how humans perceive curvature in 3D and to evaluate the ability to define a path through a 3D maze structure. Navigation through a maze is analogous to a surgeon's preoperative and intra-operative route planning, avoiding obstacles to reach targets. The *in situ* nature of grab-a-slice provides a novel way in which to study these human perceptual and cognitive processes, and compare them to those used with conventional displays.

Potential pitfalls of the device include the added space required and the ergonomics limitations on the extent of rotation and translation. The current device is also quite expensive, due to the optical tracking system, but this can be solved by substituting any of a number of cheaper technologies including other optical, RF, or inertial tracking systems, as well as mechanical encoders.

5 Conclusion

We report here the first prototype of a tracked movable display in which tomographic slices through 3D data are viewed by moving the display through the data space to show the slices *in situ*. Previously, with the SF, we employed a virtual image to produce the illusion that the surface of the patient was transparent. Here, with grab-a-slice, we use a real image to slice through what amounts to an invisible patient. Future iterations could be boom-mounted, similar to displays already used in surgical suites, or light handheld wireless devices freely manipulated in all 6 DOF. The approach holds promise for methods of graphical augmentation using stereo to improve the ability of users to understand the raw tomographic data in the context of the surrounding 3D anatomy and to improve their ability to navigate through a the 3D dataset. Finally, we are exploring the use of grab-a-slice to supervise semi-automated image analysis routines.

References

1. Levoy, M. Display of surfaces from volume data. IEEE Computer Graphics and Applications, 8, 3, 29-37 (May 1988)
2. Ware, C. and Osborne, S. Exploration and virtual camera control in virtual three dimensional environments. In Proceedings of the 1990 Symposium on interactive 3D Graphics (Snowbird, Utah, United States). ACM, New York, NY, 175-183. (1990)
3. Hinckley, K., Pausch, R., Goble, J. C., and Kassell, N. F. Passive real-world interface props for neurosurgical visualization. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Celebrating Interdependence (Boston, Massachusetts, United States, April 24 - 28, 1994). ACM, New York, NY, 452-458.
4. Fröhlich, B. and Plate, J. The cubic mouse: a new device for three-dimensional input. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (The Hague, The Netherlands, April 01 - 06, 2000). CHI '00. ACM, New York, NY, 526-531.
5. Cruz-Neira, C., Sandin, D.J., DeFanti, T.A., Kenyon, R., and Hart, J.C. The CAVE, Audio-Visual Experience Automatic Virtual Environment. Communications of the ACM, 67-72 (June 1992)
6. Browning, D.R., Cruz-Neira, C., Sandin, D.J., DeFanti T.A. Projection-Based Virtual Environments and Disability. Virtual Reality Conference, 1994.
7. Looser, J., Billinghurst, M., and Cockburn, A. Through the looking glass: the use of lenses as an interface tool for Augmented Reality interfaces. In Proceedings of the 2nd international Conference on Computer Graphics and interactive Techniques in Australasia and South East Asia (Singapore, June 15 - 18, 2004). S. N. Spencer, Ed. GRAPHITE '04. ACM, New York, NY, 204-211.

8. Kalkofen D., Mendez E., Schmalstieg D. Interactive Focus and Context Visualization for Augmented Reality. In Proceedings of 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, Nara, Japan, 191-200 (Nov 2007)
9. State, A., Livingston, M. A., Garrett, W. F., Hirota, G., Whitton, M. C., Pisano, E. D., and Fuchs, H. Technologies for augmented reality systems: realizing ultrasound-guided needle biopsies. In Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques SIGGRAPH '96. New York, NY, 439-446.
10. Sauer F., Khamene A., Bascle B., Schimmang L., Wenzel F., Vogt S. Augmented Reality Visualization of Ultrasound Images: System Description, Calibration, and Features. IEEE and ACM International Symposium on Augmented Reality. (2001)
11. Stetten, G. , Chib, V. Overlaying Ultrasound Images on Direct Vision. Journal of Ultrasound in Medicine, 20, 3, 235-240 (2001)
12. Hofstein, S. Ultrasonic Scope. US Patent no. 4,200,885, Apr. 1980.
13. Masamune, K., Masutani, Y., Nakajima, S., Sakuma, I., Dohi, T., Iseki, H., and Takakura, K. Three-Dimensional Slice Image Overlay System with Accurate Depth Perception for Surgery. In Proceedings of the Third international Conference on Medical Image Computing and Computer-Assisted intervention (October 11 - 14, 2000). S. L. Delp, A. M. DiGioia, and B. Jaramaz, Eds. Lecture Notes In Computer Science, vol. 1935. Springer-Verlag, London, 395-402.
14. Masamune, K., Fichtinger, G., Deguet, A., Matsuka, D., and Taylor, R. 2002. An Image Overlay System with Enhanced Reality for Percutaneous Therapy Performed Inside CT Scanner. In Proceedings of the 5th international Conference on Medical Image Computing and Computer-Assisted intervention-Part II (September 25 - 28, 2002). T. Dohi and R. Kikins, Eds. Lecture Notes In Computer Science, vol. 2489. Springer-Verlag, London, 77-84.
15. Masamune, K., Sato, I., Liao, H., and Dohi, T. 2008. Non-metal Slice Image Overlay Display System Used Inside the Open Type MRI. In Proceedings of the 4th international Workshop on Medical Imaging and Augmented Reality (Tokyo, Japan, August 01 - 02, 2008). T. Dohi, I. Sakuma, and H. Liao, Eds. Lecture Notes In Computer Science, vol. 5128. Springer-Verlag, Berlin, Heidelberg, 385-392.
16. Chang, W., Amesur, N., Klatzky, R., Zajko, A. , and Stetten, G. Vascular Access: Comparison of US Guidance with the Sonic Flashlight and Conventional US in Phantoms. Radiology, 241, 771-779 (Dec 2006)
17. Wu, B., Klatzky, R., Shelton, D., Stetten, G., Psychophysical Evaluation of In situ Ultrasound Visualization IEEE Transactions on Visualization and Computer Graphics (TVCG), Special Issue on Haptics, Virtual and Augmented Reality, 11, 6, 684-693 (Nov/Dec 2005)
18. Klatzky, R., Wu, B., Shelton, D., Stetten, G. Learning to perform actions in near space under perceptual vs. cognitive control. ACM Transactions on Applied Perception. Vol. 5, No. 1 (Jan 2008)
19. Wu, B., Klatzky, R., Shelton, D., Stetten, G. Mental concatenation of perceptually and cognitively specified depth to represent locations in near space. Experimental Brain Research, 184, 3, 295-305 (2008)
20. Klatzky, R., Wu, B., Stetten, G. Spatial representations from perception and cognitive mediation: The case of ultrasound. Current Directions in Psychological Science (in press).
21. Shelton, D. Virtual Tomographic Reflection for Psychophysical Analysis of the Sonic Flashlight. Ph.D. Dissertation, Robotics Institute, Carnegie Mellon University, advisor G. Stetten, March 2007.
22. Watt, A. Fundamentals of Three-Dimensional Computer Graphics. Reading, MA: Addison-Wesley (1990)

Session 4 – Visualisation and Algorithms

Patient-specific Texture Blending on Surfaces of Arbitrary Topology

Johannes Totz, Adrian J. Chung, Guang-Zhong Yang

Imperial College London

Abstract. For surgical simulation, generating seamless texture maps from video examination data is an important step of high-fidelity subject-specific model creation both in terms of dynamic morphological representation and visual appearance. Existing techniques have encountered a number of difficulties including handling of surface topology, visual fidelity, and algorithmic stability. To tackle these problems, we propose a robust texture blending algorithm for complex surfaces with arbitrary topology. A patient-specific organ mesh is parameterised patch-wise and textured with video image sequences matched with the 3D tomographic model. To reduce seams in the texture caused by colour and intensity variations, a multi-band blending approach is adopted. Difficulties associated with filtering in the discontinuous texture domain are avoided by using an overlapping parameterisation scheme. Detailed user evaluation demonstrates the strength and practical value of the proposed method.

Keywords: Multi-band blending, seam reduction, patient-specific, patch-wise parameterisation

1 Introduction

Due to the high degrees of manual dexterity involved in Minimally Invasive Surgery (MIS), the use of simulators with standardised or idealised patient anatomy has played an important role in basic surgical training and skills assessment [1]. Further advances in the field have called for the need of incorporating patient-specific models for high-fidelity simulation [2] such that they can be used for advanced training on a wide variety of real patient data illustrating both natural diversity and particular pathognomonic cases. For endoscopic simulation, for example, instead of synthesizing surface texture from a generic procedural model, it is possible to combine 3D tomographic data with actual subject-specific endoscopic video for realistic simulation. Key technical issues associated with this approach are related to 2D/3D registration, handling dis-occlusion artefacts, and view invariant surface texture extraction and mapping. For the latter, artefacts can arise due to inconsistent illumination caused by the moving light source and mis-registration errors between the video frame and 3D model. Furthermore, different video frames in the endoscope sequence can be mapped to the same surface area; any misalignment between them can cause a noticeable seam. In practice, even if features are perfectly aligned, a visible seam can also result from differences in brightness caused by varying gains. Although surface-oriented blending operations can alleviate this problem to some extent, over-blurring and ghosting artefacts can often result. Furthermore, the non-trivial topology of human anatomical structure can complicate the application of texture blending operations, which have traditionally been limited to planar rectilinear domains.

The problem stated above is closely related to image mosaicing and seam reduction, which is a well studied field for two-dimensional photography. Burt and Adelson [3] presented the idea of blending overlapping images at different frequencies in order to preserve image details that would be blurred otherwise. Atasoy et al. used planar domain blending for automatic fibroscopic video mosaicing in minimally invasive surgery [4]. Extending the multi-band approach to non-planar domains is not trivial as there is no general method for mapping meshes of arbitrary topology to the planar domain without introducing severe distortions and discontinuities. In the field of 3D object scanning, Baumberg [5] presented a method to generate weight maps for multi-band blending. This was extended by Allene et al. [6] by including optimised seam-placement [7] and blending via feathering. For virtual bronchoscopy, Chung et al. [8] presented a method to increase the quality of texture maps by estimating the BRDF of the bronchial wall and factorising this into illumination-independent textures for a given video image. While subject specific-textures were then blended based on the intensity predicted by a simplified BRDF model increasing the quality, the actual blending process did not take feature sizes into account.

The purpose of this paper is to propose a method for constructing a patient specific appearance model that can be applied to any arbitrary manifold surface, thus catering for the wide variety of topologies found in human anatomy. A piecewise surface parameterisation is automatically constructed that guarantees algorithmic stability for arbitrary surface meshes, thus allowing image mosaicing to be extended to generalised domains with seamless reconstructed tissue surface textures.

2 Method

To illustrate the practical clinical application of the proposed technique, we will use bronchoscopy as an example to outline the key technical steps involved. Consenting bronchoscopy patients were scanned during a single breathhold using a Siemens Somatom Volume Zoom 4-Channel Multidetector CT with a slice thickness of 3 mm and in-plane resolution of 1 mm. The patients subsequently underwent a bronchial examination with an Olympus BF-type (120 degree FOV) bronchoscope which was recorded on PAL DV tape. The CT data was segmented into bronchial airways and a surface triangle mesh was extracted with Marching Cubes. The surface mesh and the video frames are registered using a *pq*-space method [9]. The main steps in our method for texture blending are depicted in fig. 1.

2.1 Parameterisation

The surface mesh was parameterised to form the texture domain (fig. 1a). First, it was cut into disc-homeomorphic patches by a region growing algorithm, seeded from random vertices so that the triangles of each region approximate the Voronoi con-

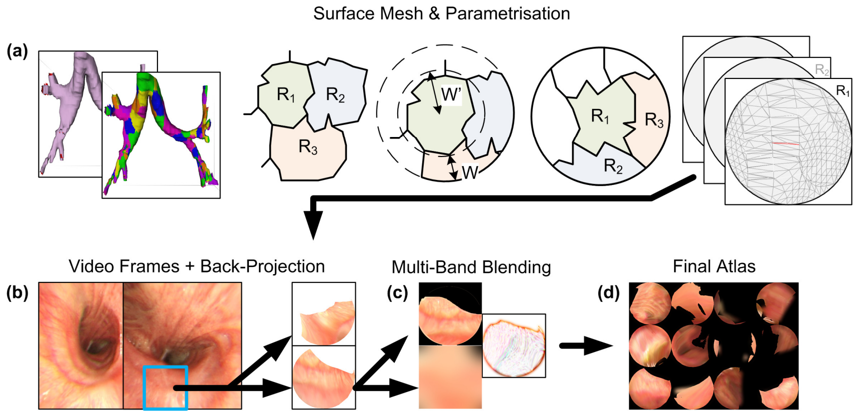


Fig. 1. Illustration of the key steps in our texture blending framework. **(a)** The input mesh was cut into individual patches, each geometrically dilated so that $W \approx W'$ to overlap with its neighbours and parameterised into the plane using mean-value-coordinates. **(b)** Video frames were back-projected into the texture space. **(c)** Within each texture patch multi-band blending was performed to smoothen out the colour and intensity differences of the input video frames, while preserving detail features. **(d)** The final atlas is generated by copying all patches into one shared texture.

straint on the surface of the mesh similar to [10]. The distance measure used here was the edge-length of the dual graph of the mesh. Each grown region was subject to consistency checks that ensured disc-homeomorphism, i.e., (1) the region's perimeter should not self-intersect; (2) the perimeter should consist of one connected component only; (3) neighbouring regions should share perimeter in one contiguous chain; and (4) all triangles of a region must form a genus-0 surface. These checks in combination with the Voronoi constraint minimise distortion in the subsequent parameterisation.

Each region was then geometrically dilated, while preserving genus-0, to include triangles from its neighbouring regions, forming the set of patches. The number of triangles added to the region R_i to form a patch was defined so that the shortest distance W between the parameterisation border (in 3D space) and any triangle of R_i approached the radius of R_i (cf. fig. 1a). This dilation is an important step for the actual blending procedure. Each patch was parameterised into the plane using mean-value-coordinates [11] which guarantees crucial injectivity (no self-overlap) as long as the boundary is convex, in our case circular. Numerically, this involves solving two sparse systems per patch, one for the u-coordinate and one for v, which established a discontinuous texture domain for the whole surface. The system is defined by (after [12]):

$$w_{ij} = \frac{\tan(\delta_{ij}/2) + \tan(\gamma_{ij}/2)}{||v_j - v_i||} \qquad \lambda_{ij} = w_{ij} \Big/ \sum_{k \in N_i} w_{ik} \qquad f(v_i) = \sum_{j \in N_i} \lambda_{ij} f(v_j) \tag{1}$$

where δ and γ are the angles between the edges of adjacent vertices, thus minimising the distortion induced.

Video images were then back-projected onto the mesh and into the texture space. This yielded a set of patches that were visible from the current camera position and each patch had its pixel content back-projected, with parts being shared with its neighbours.

As the camera moved inside the bronchial lumen, the back-projected pixel content could change in colour, features and illumination. All back-projections were therefore required to be blended to form a single texture map.

2.2 Blending

Simply blending the pixel content from different frames can cause noticeable artefacts as mentioned earlier due to differences in illumination and (unavoidable) mis-registration. Instead, multi-band blending blurs pixel content across different scales [3]. That is, low frequency content is blended over a long range, while high frequency details are blended over a short range.

In this study, input images were decomposed into frequency bands via the Laplacian image pyramid and blended with weight-maps that define the seams between images (feathering). The pixel content of different video frames projected into the parameterised patches of the mesh was blended in the texture domain. In general one would need to pay special attention to border conditions for the Gaussian filter; but the most difficult part is to work around the discontinuity of the texture map: neighbouring triangles were not necessarily neighbours in the texture map and therefore created a seam. If blended as-is, these seams showed up in final renderings as shown in fig. 2c. Instead, the overlap of texture patches introduced during the parameterisation simplified the filtering operation significantly. In this way, the filter had access to pixel content from neighbouring triangles and patches and no special border conditions were necessary.

For this processing step, the weight maps were derived from each texel's distance to the camera and angle between surface normal and view direction. This is similar to the "best-camera" approach taken in previous methods that acquire texture maps for 3D-scanned objects. For every texture patch the weight maps from all back-projections were binary-ised to determine the colour texel that was "most responsible" for every texel in the final texture map. These binary weight maps were successively smoothed, along with the back-projected colour texels.

In addition, a "base colour" patch was synthesised from all input to fill in blank texels. This base colour patch contributed to the blending so that the weights for any given texel always summed to 1. It was generated by successively copying all input images into the base-colour patch and smoothing it with decreasing filter width. This caused it to have the overall colour appearance of the input images. The blending process can be summarised by:

$$R_i = \sum_{j=1}^{b} \left((F_{i,j+1} \circ G_{\sigma,j+1}) - (F_{i,j} \circ G_{\sigma,j}) \right) \cdot (M_{i,j} \circ G_{\sigma,j}) \tag{2}$$

where $F_{i,j}$ is the texture patch for frame i at level j of the Gaussian image pyramid, $M_{i,j}$ the binary-ised weight map, $G_{\sigma,j}$ the Gaussian kernel with standard deviation σ, and R_i the final texture patch blended over b frequency bands.

The parameterisation assigned *uv*-coordinates to vertices of the mesh which needed to be scaled to image dimensions (256 by 256 pixels in this case), therefore requiring the pixel width of the filter to be set appropriately. The initial filter kernel is sized according to the average area change ratio of the triangles of the patch induced by the parameterisation and *W*. This corrects for the scale-mismatch of the fixed-sized image dimensions of the texture patch with respect to the actual surface area covered on the mesh.

Individual texture patches were packed into an atlas and rendered with OpenGL to produce views that were not in the original footage. From every texture patch only the texels that were actually mapped onto the region grown earlier, were used for rendering. Because every patch contained texels from a region's neighbours the blending operation produced continuous results for patches that were otherwise discontinuous in texture space.

3 Results

3.1 Bronchoscopy Video

Examples of blending results are shown in fig. 2. It is evident that the proposed method is clearly superior; simple stitching of textures without any blending produces most obvious artefacts whereas alpha-blending has the problem that, in order to achieve noticeable seam reduction the filter kernel needs to be large so as to blur out colour and illumination differences, which in turn causes excessive blurring of small high frequency details. Multi-band blending (figs. 2c and d) promises to alleviate this by blending features over a different range based on what frequency band they are in.

Critical is the region of overlap of the parameterisations since the texture domain will be discontinuous but the filter operation requires access to similar pixels in adjacent texture patches. If this is not accounted for, intensity and colour differences are not handled, as can be seen in fig. 2c. The parameterisation in section 2.1 facilitates this by not just providing an overlap but also sizing it appropriately and consistently for the varying scales of the filtering operation in the multi-band blending.

3.2 Validation

To evaluate the visual quality of the proposed scheme, a user experiment was carried out. A total of 13 subjects (medical imaging researchers) were recruited and asked to select one image they regard had the least artefacts, without participants being conditioned as to what artefacts to look for, i.e. the meaning of artefact was explicitly not defined. Thus the results of the survey will determine whether the subjects considered seams to be major artefacts in texture maps. Participants were not presented real bronchoscopy footage with which to compare because the focus of this research is not to evaluate realistic simulation of illumination and shading.

Images were presented four per slide showing the same view (but different blending) from a randomised pool, resulting in a total of ten slides. Fig. 3a shows the overall scores derived. Multi-band blending is clearly preferred by most subjects (64%), followed by alpha-blending (23%). Several subjects preferred multi-band blending without parameterisation overlap. Although this method blends features accurately inside a region, there is significant intensity mismatch at the borders (fig. 2c fourth column), an artefact that the parameterisation presented above helps minimising. Only two subjects found simplistic stitching (cf. fig. 2a) to be superior in three cases. In addition, the participants were asked for the "second best" image. If multi-band blending was found to not be the "best" image it was still rated second best for 88% of the images.

One advantage of multi-band blending is that it smoothens out brightness differences. This manifests itself in a lower standard-deviation of the overall brightness. To measure this effect the standard deviation of the brightness was cal-

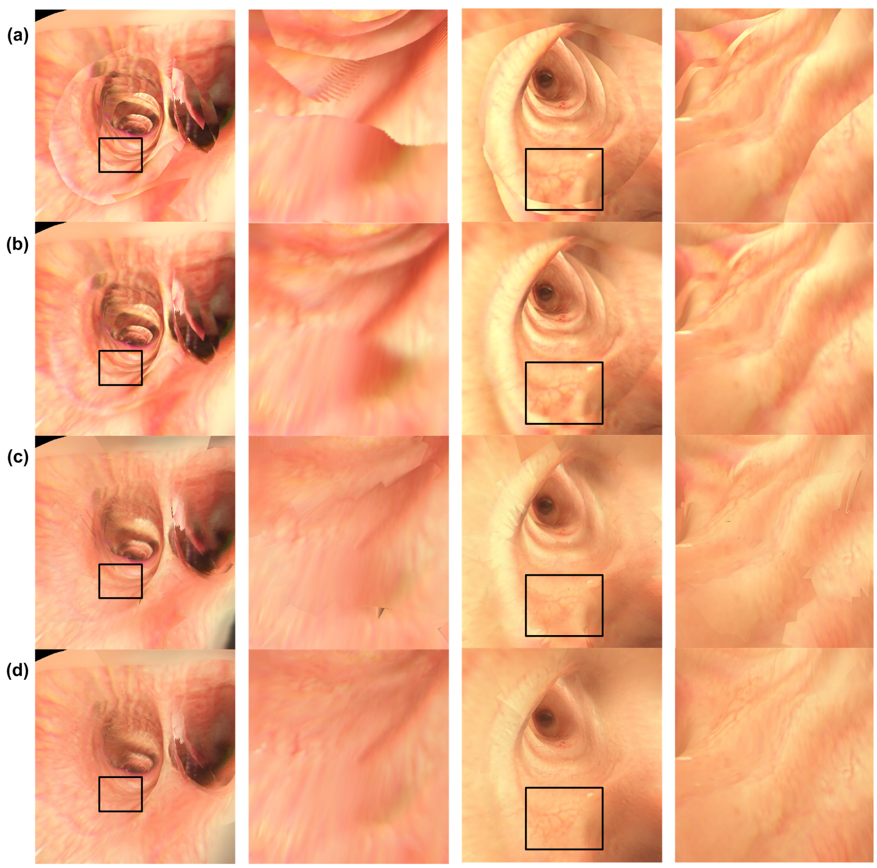


Fig. 2. Different texture blending methods for a virtual bronchoscopic view. This shows two patient-specific data sets rendered to produce views not in the original examination footage. The blending methods compared are: **(a)** simple stitching, no blending; **(b)** simple alpha-blending; **(c)** Multi-band blending without parameterisation overlap. **(d)** Multi-band blending using our method.

culated for different methods by sampling it evenly across the textured mesh. Fig. 3b shows a plot of the standard deviation relative to the simple stitching method. As stitching yields the strongest artefacts with large brightness shift it was chosen as the baseline to compare the other methods with. Multi-band blending, as proposed in this paper, has an average standard deviation compared to stitching of 71% and 85%, for two patients respectively, meaning that textures had a more consistent brightness than with the other methods. However, because the brightness is blended over a long range, this can introduce an error by altering the average brightness of the texture map. To assess this effect in detail, the input brightness was calculated and compared to the output brightness. A slight tendency towards brightening was observed for our multi-band approach, with an average brightness increase of 8% and 7% for two patients, respectively. This is caused by blending video frames that are taken close to the organ's surface with frames taken further away from it. The correct way to account for these frames is to generate an illumination-independent texture map before blending, for example as in [8].

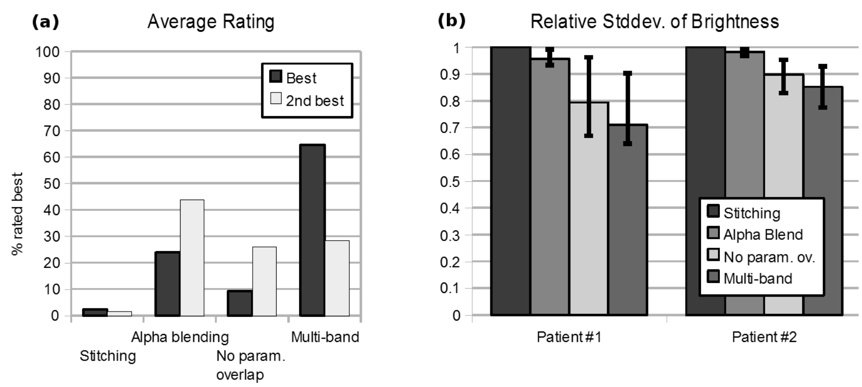


Fig. 3. **(a)** User ratings for the different blending methods (cf. fig. 2). The average per-method rating is shown, with multi-band blending scoring 64% for "best looking" and 28% for "second best". Alpha blending scored 23% as it is able to reduce seams, at the cost of blurred features though. "No param. ov." refers to multi-band blending without a parameterisation that includes overlapping regions. **(b)** Standard deviation of brightness after blending, relative to stitching. Our multi-band blending approach has the lowest brightness deviation, meaning that resulting textures have consistent brightness, despite a small bias towards brightness increase (cf. text).

4 Discussion

In this paper, we have presented a method for performing stable multi-band blending on arbitrary surface meshes. The key is a surface parameterisation that is able to provide pixel neighbourhood to ensure continuity in a globally discontinuous domain. After the organ of interest has been segmented and converted into a surface mesh, only limited user interaction is required to sparsely select video frames (to reduce computational cost) and exclude sequences with large artefact (e.g. coughing). The method is inherently parallelisable and amenable to GPU implementation, allowing real-time application. Given that the registration method used to co-register video and mesh performs reasonably well, the pre-operatively obtained mesh can be textured intra-operatively, not just for bronchoscopy but also other endoscopic procedures, thus opening a wide range of applications beyond training and simulation.

References

1. Seymour, N.E., Gallagher, A.G., Roman, S.A., O'Brien, M.K., Bansal, V.K., Andersen, D.K., Satava, R.M.: Virtual reality training improves operating room performance. Annals of Surgery **236**(4) (October 2002) 458–464
2. Mutter, D., Bouras, G., Marescaux, J.: Digital technologies and quality improvement in cancer surgery. European Journal of Surgical Oncology **31**(6) (August 2005) 689–694
3. Burt, P.J., Adelson, E.H.: A multiresolution spline with application to image mosaics. ACM Transactions on Graphics **2**(4) (October 1983) 217–236
4. Atasoy, S., Noonan, D.P., Benhimane, S., Navab, N., Yang, G.Z.: A global approach for automatic fibroscopic video mosaicing in minimally invasive diagnosis. In Metaxas, D., Axel, L., Fichtinger, G., Székely, G., eds.: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008. Volume 5241/2008 of Lecture Notes in Computer Science., Springer Berlin/Heidelberg (2008) 850–857
5. Baumberg, A.: Blending images for texturing 3d models. In: Proc. Conf. on British Machine Vision Association. (2002) 404–413
6. Allène, C., Pons, J.P., Keriven, R.: Seamless image-based texture atlases using multi-band blending. In: 19th International Conference on Pattern Recognition. (2008)
7. Lempitsky, V., Ivanov, D.: Seamless mosaicing of image-based texture maps. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR '07. (June 2007) 1–6
8. Chung, A.J., Deligianni, F., Shah, P., Wells, A., Yang, G.Z.: Patient-specific bronchoscopy visualization through brdf estimation and disocclusion correction. IEEE Transactions on Medical Imaging **25**(4) (April 2006) 503–513
9. Deligianni, F., Chung, A., Yang, G.Z.: *pq*-space based 2d/3d registration for endoscope tracking. In Ellis, R.E., Peters, T.M., eds.: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2003. Volume 2878/2003 of Lecture Notes in Computer Science., Springer Berlin/Heidelberg (2003) 311–318
10. Eck, M., DeRose, T., Duchamp, T., Hoppe, H., Lounsbery, M., Stuetzle, W.: Multiresolution analysis of arbitrary meshes. In: SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques, New York, NY, USA, ACM (1995) 173–182
11. Floater, M.S.: Mean value coordinates. Computer Aided Geometric Design **20**(1) (2003) 19–27
12. Floater, M.S., Hormann, K.: Surface Parameterization: a Tutorial and Survey. Mathematics and Visualization. In: Advances in Multiresolution for Geometric Modelling. Springer Berlin Heidelberg (2005) 157–186

Improving Tracking Accuracy Using Neural Networks in Augmented Reality Environments

Arun Kumar Raj Voruganti, Oliver Burgert, and Dirk Bartz

Innovation Center Computer Assisted Surgery (ICCAS),
Universität Leipzig, Germany. *

Abstract. Integration of Augmented Reality (AR) systems in medicine requires good accuracy. Current planar-marker based AR system accuracy is limited by the techniques used such as camera calibration, pose estimation, inter-space calibration. Accuracy analysis of individual techniques and reports on dispersion of accuracy is necessary to estimate overall accuracy of AR for medical applications. In this work, we discuss our observation on the deviation of actual to estimated distance due to distortion of marker features in planar-marker tracking. Based on parameters of marker features, we have defined a model of the tracking deviation. We discuss the dependency of tracking deviation on these pseudo-parameters. To improve tracking accuracy, we proposed usage of neural networks to predict tracking deviation. Using this predicted tracking deviation, we show in results that the estimated distance can be corrected and that the overall tracking accuracy can be improved.

1 Introduction

Augmented Reality (AR) is a promising technology for the integration of patient specific 3D image data into the operating view of a surgeon. In computer-assisted surgery, the objective of an AR system is to register and superimpose patient specific data over a continuous acquired image stream. With this enhanced vision, orientation problems during image-guided navigation through the patient anatomy can be solved. Such an enhanced vision can be used in surgical assistance, surgical training and surgical simulation. Fig. 1 shows various steps of AR work flow. Numerous research has been carried at finding better methods for camera calibration [1, 2], inter-space calibration [3, 4], pose estimation [5–8] and advanced in-situ visualization [9]. A formal model and description of frequent occurring patterns in AR modeling are given in [10].

Although AR technology is established in some fields, one of the main limitations of AR technology in medical applications is its reliability. To incorporate AR technology in medical applications, sufficient accuracy of the system is necessary. Currently AR systems lack accuracy due to the errors propagated in

* ICCAS is funded by the German Federal Ministry of Education and Research and the Saxon Ministry of Science and the Fine Arts with grant numbers 03 ZIK 031 and 03 ZIK 032.

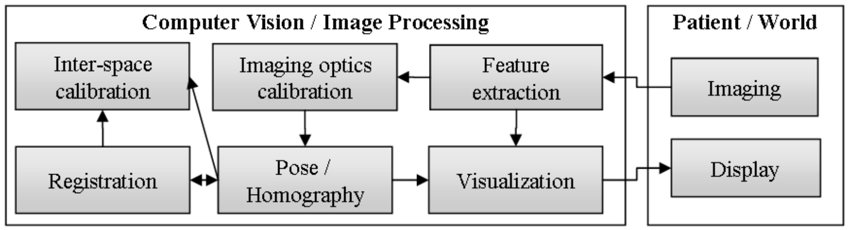


Fig. 1. Augmented reality work flow components.

the AR work flow. Thus a thorough analysis of error sources has to be done in the AR work flow. Also it is important to identify the factors influencing these errors. There have been attempts to estimate errors and improve accuracy in AR systems. Tracking is seen as the most crucial part that affects the overall accuracy of the AR systems. Systematic tracking errors occurring due to sensor distance, angle and size of the marker in marker based tracking have been studied in [11, 12]. Methods to predict such systematic errors based on the marker features have been proposed in [13, 14]. Using multi-markers, improvement in tracking accuracy has been noticed, but this requires large multi-marker configuration [15].

In this work, we developed an AR system based on a planar marker tracking technique. The main source of error in such an AR system is due to tracking. We report our observation on tracking deviation, and propose a method based on neural networks to reduce this error and to improve overall tracking accuracy.

2 Methods

2.1 Error sources in AR

We define the error of an AR system as the deviation of distance between the estimated position to its corresponding real world position in a common reference coordinate system. This error usually occurs due to the propagated errors from different components of an AR system shown in Fig. 1. We categorize errors into two types; dynamic and static. Camera calibration gives the projection matrix and the distortion parameters. Although precise calibration is not possible with a planar calibration body and results vary for every set of images, it is a common practice to calibrate an optical system a number of times and take an average until a stable result is acquired. Since this calibration errror does not change through-out the AR workflow, we denote this error as static error. In contrast, changing errors are denoted as dynamic error.

2.2 Parameterizing tracking accuracy

Tracking with planar markers involve segmentation of features from video images and pose estimation of the planar object in camera coordinates. The precision of segmentation is dependent on the illumination and the underlying algorithms. Accuracy of the pose estimation is dependent on the marker size, distance at which the marker is placed from camera, and the camera view angle with respect to the plane of the planar object. Also the integrity of projection matrix and focal length acquired in camera-calibration step are important for tracking, where focal length is used in pose estimation algorithms.

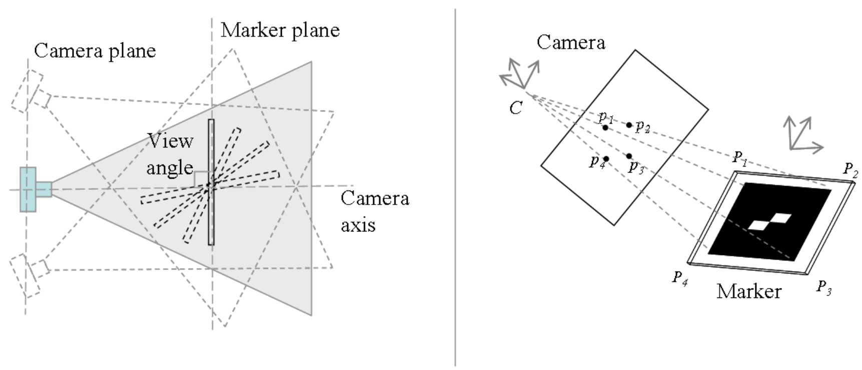


Fig. 2. (left) Geometry of perspective distortion, (right) Perspective projection of planar marker.

Fig. 2 shows the geometry of perspective distortion and perspective projection of a planar marker. Perspective distortion is caused when the marker plane is not parallel to the camera plane. When the marker plane is parallel to the camera plane, marker corners used for pose estimation are not distorted, and thus pose estimation of the planar object is accurate. The more inclined the object plane with respect to the camera plane is, the greater is the distortion of object features and the inaccuracy of pose estimation.

We define here parameters that are used to parameterize calculated deviation. These parameters are calculated from the features of the planar marker. Let's say $p_1 ... p_4$ are the tracked features (corners) of the planar-marker pattern shown in Fig. 2(right) and c is the center of the marker. Pixel area of the marker region, variance of distance from center to corner and mean-intensity can be calculated as follows:

$$A = \frac{d_{p_1 p_3} \times d_{p_2 p_4}}{2} \tag{1}$$

$$V_d = \frac{\sum_{i=1}^{4} \left(d_i - \overline{d}\right)^2}{4} \tag{2}$$

$$I = \frac{1}{N} \sum_{x=1}^{N} \sum_{y=1}^{N} f(x, y) \quad (3)$$

Where $d_{p_1p_3}$ and $d_{p_2p_4}$ are the length of diagonals of the marker square, $d_1...d_4$ are the distances from marker-center to corners $p_1...p_4$ and $\overline{d}$ is the mean of the center to corner distances. N is the total number of pixels in an image and $f(x, y)$ is the intensity of a single pixel at (x, y).

Having these parameters defined, we can study the deviation of estimated position to the actual deviation. Both, the pixel area and variance, vary according to the distance and the angle at which the marker is placed with respect to the camera. Let us assume that the marker is placed at a fixed distance from the camera center and can be rotated around the X-Y axis about its center between two pivots as shown in Fig. 3 (left).

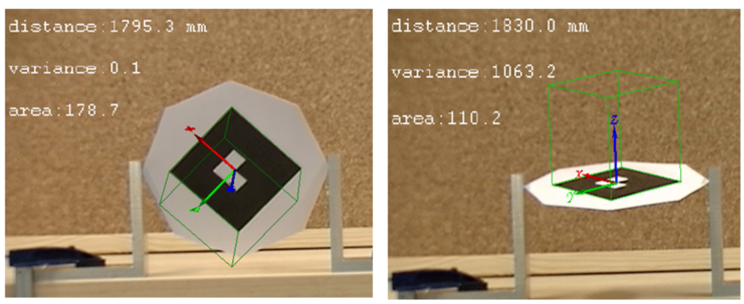


Fig. 3. Estimated distance of the marker fixed between pivots to the camera, detected pixel area (black region) and variance of center to corner length.

Now if the marker plane (X-Y) is perfectly parallel to the camera plane, the marker is fully visible and the detected features are not distorted. The pixel area of the detected region should be at its maximum. As the marker is inclined, the area would decrease depending on the inclination of X-Y plane to the camera plane. This can be observed in Fig. 3 (right). The variance on the other hand would be zero if the marker X-Y plane is parallel to the camera plane and increases as the angle of inclination of the X-Y plane increases with respect to the camera plane as shown in the Fig. 3 (right). From the images above, it can also be noted that the distance calculated from the estimated pose varies according to the inclination of the marker X-Y plane with the camera plane. The more the marker is inclined to the camera plane, the greater is the influence of distortions and the lesser is the accuracy of the pose estimation. In other words, the greater the variance of marker-center to corner length, the greater is the perspective distortion and the lesser is the accuracy of the estimated pose. Tracking deviation also depends on various other dynamic factors such as illumination, segmentation and angle of inclination. We assume that the sementation of features is dependent on the illumination and so the tracking deviation. We consider here mean intensity as the parameter measuring surrounding illumination that influences the tracking deviation.

2.3 Modeling tracking deviation with neural networks

We propose a method to improve tracking accuracy by application of neural networks. Neural networks are information processing systems with interconnected artificial neurons. Neural networks use computation models to process data that mimics biological nervous system. Neural networks can be basically thought of as a black box capable of learning non-linear mathematical functions. Neural networks have been successfully applied in various fields of science for function approximation, classification and clustering problems [16]. Here, we use neural networks to approximate the tracking deviation. Fig. 4 depicts the model of the network used for function approximation.

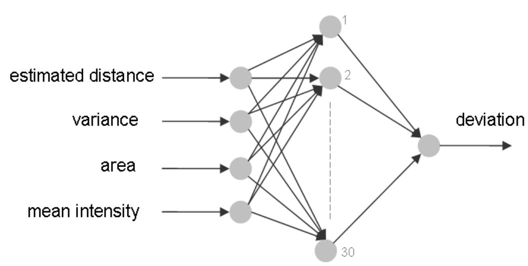


Fig. 4. Three-layer neural network model.

We used a three layered neural network topology with 4 input neurons, 30 hidden neurons and 1 output neuron. The input to the network are the marker-to-camera-distances calculated from the estimated pose, pixel area, variance of marker-center to corner distance and mean-intensity calculated from the video image. Output of the network is the deviation of the estimated distance from the actual distance. We take into consideration actual or most accurate distance as the one calculated at minimum distortions when the marker plane is perfectly parallel to the camera plane.

We have collected different training and testing dataset by positioning the marker at varying distances from the camera. The inclination of the marker to the camera was varied during the course of the data collection. The network was trained with standard back propagation algorithm for around 1000 epochs. All the parameters were normalized prior to training with the minimum and maximum values collected for each parameter in the dataset.Once the network is trained with the training data set, the saved network can be used to test the testing dataset. For every given input, the network would predict the deviation of the estimated position to the actual position. With this prediction, tracking deviation is known and can be used to correct the estimated position to improve tracking accuracy. Subtracting the predicted deviation from the estimated position would then give the corrected or actual position of the marker to camera.

3 Results

Table 1 lists statistics for deviation of estimated distance (marker to camera) before correction, and deviation of estimated distance after correction with the neural network predicted value. Column one and two lists the distance at which the marker is placed from camera, and total number of positions collected for each distance.At a first glance, it can be noted that the deviation increases as the marker distance to camera increases. Also, the dispersion of the deviation from its mean increases as the marker distance to camera increases. Comparing the deviation statistics before and after correction, it is observed that the deviation after correction has been significantly reduced. This means improvement in tracking accuracy after correction of estimated distance with the neural network predicted deviation.

Table 1. Tracking deviation statistics

		Deviation of estimated marker distance to camera							
		Before correction(mm)				After Correction(mm)			
distance (mm)	total	max	min	mean	std	max	min	mean	std
997.9	421	13.7	−2.4	4.282	3.707	3.891	−5.423	0.618	1.725
1198.85	473	16.85	−4.65	3.522	5.602	1.2	−4.718	−1.037	1.079
1400.7	714	23.4	−4	5.875	7.569	2.689	−3.0625	0.416	1.222
1500.6	703	25.5	−6.3	1.376	5.895	5.965	−4.831	0.461	2.125
1600.1	499	31.3	−7	3.909	9.296	4.268	−5.069	0.054	2.037
1707.6	1007	36.2	−8.4	2.998	10.339	6.252	−6.207	−0.539	2.688
1799.8	582	41.2	−0.5	11.734	12.412	8.495	−4.736	0.890	2.826

The neural network prediction is not accurate to the actual deviation as there exists deviation still after correction of the estimated position with the predicted value. But it can be inferred from the results, that it is possible to reduce tracking deviation error using neural network prediction signifantly. Comparing the standard deviation before and after correction, it can be noticed that the neural networks could predict well within a range according to the input image features discussed in Section 2. Knowing this deviation, and dispersion range from its mean, one can create accuracy reports for specific tracking methods. Fig. 5 shows graphs of the estimated marker to camera distance when the marker is placed at distances 1400.7mm and 1799.8mm. The inclination of the marker plane to camera plane was varied during the collection of data. X-axis on the plots shows the collection number for the position and Y-axis shows the distance

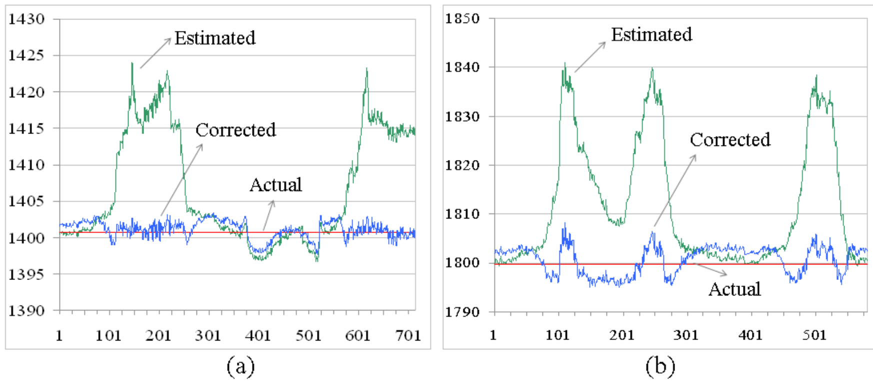


Fig. 5. Marker to camera distance before (green) and after correction (blue) at 1400.7 mm (a) and at at 1799.8 mm (b).

calculated from to the camera. The red line in both plots is the actual or accurate distance of the marker to the camera, green line-plot is the estimated distance and the blue line-plot is the distance after correction. From both the plots, it can be observed that the estimated distance disperses away from the actual in certain regions. This tracking inaccuracy is due to increasing inclination of the marker plane to the camera plane. From the blue line-plot it is observed that the deviation has reduced after correction with the neural network prediction, and the dispersion about the actual distance is also less compared to green line-plot.

4 Conclusion

We discussed how tracking accuracy is influenced with the inclination of marker plane to the camera plane. Based on our observations, we have defined pseudo-parameters calculated from image features, that can be used to study tracking deviation and proposed a method to correct the deviation by using neural network prediction. Neural network prediction of deviation showed good results in tracking accuracy improvement. The neural network model as proposed in Section 2 can be extended with other input parameters that can influence tracking accuracy. It should be noted that machine learning methods for improving tracking accuracy requires good model and dataset collection for training and testing.

As a continuation of our research, we would like to extend the parameters and neural network model for better prediction of the tracking deviation. By having a precise setup, we can overcome the assumption we made in this work, that the actual or most precise distance is the one estimated when the marker plane is parallel to the camera plane. Although this assumption is true and is also clearly seen from the plots in Fig. 5, we can also include the angle of inclination of marker to the camera plane in the neural network model with an accurate experimental setup. The proposed method can be used in AR based range measurement applications.

References

1. Zhang, Z., Zhang, Z.: A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (1998) 1330–1334
2. Wengert, C., Reeff, M., Cattin, P.C., SzékelyProc, G.: Fully automatic endoscope calibration for intraoperative use. In: Proc. of Bildverarbeitung für die Medizin. (2006) 419–23
3. Park, F., Martin, B.: Robot sensor calibration: Solving Ax=Xb on the euclidean group. IEEE Transactions on Robotics and Automation **10** (1994) 717–721
4. Tsai, R., Lenz, R.: A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. IEEE Transactions on Robotics and Automation **5** (1989) 345–358
5. Dementhon, D.F., Davis, L.S.: Model-based object pose in 25 lines of code. International Journal of Computer Vision **15** (1995) 123–141
6. Horn, B.K.P., Hilden, H.M., Negahdaripour, S.: Closed-form solution of absolute orientation using orthonormal matrices. Journal of the Optical Society of America **5** (1988) 1127
7. Lowe, D.G.: Fitting parameterized three-dimensional models to images. IEEE Transactions on Pattern Analysis and Machine Intelligence **13** (1991) 441–450
8. Quan, L., Lan, Z.: Linear n-point camera pose determination. IEEE Transactions on Pattern Analysis and Machine Intelligence **21** (1999) 774–780
9. Bichlmeier, C., Wimmer, F., Heining, S.M., Navab, N.: Contextual anatomic mimesis: Hybrid in-situ visualization method for improving multi-sensory depth perception in medical augmented reality. In: Proc. of the IEEE and ACM International Symposium on Mixed and Augmented Reality. (2007) 129–138
10. Pustka, D., Huber, M., Bauer, M., Klinker, G.: Spatial relationship patterns: elements of reusable tracking and calibration systems. In: Proc. of the IEEE and ACM International Symposium on Mixed and Augmented Reality. (2006) 88–97
11. Abawi, D.F., Bienwald, J., Dorner, R.: Accuracy in optical tracking with fiducial markers: An accuracy function for ARToolKit. In: Proc. of IEEE and ACM International Symposium on Mixed and Augmented Reality. (2004) 260–261
12. Kato, H., Billinghurst, M.: Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In: Proc. of IEEE and ACM International Workshop on Augmented Reality. (1999) 85
13. Bauer, M., Schlegel, M., Pustka, D., Navab, N., Klinker, G.: Predicting and estimating the accuracy of n-occular optical tracking systems. In: Proc. of IEEE and ACM International Symposium on Mixed and Augmented Reality. (2006) 43–51
14. Davis, L., Clarkson, E., Rolland, J.P.: Predicting accuracy in pose estimation for marker-based tracking. In: Proc. of IEEE and ACM International Symposium on Mixed and Augmented Reality. (2003) 28
15. Yoon, J.H., Park, J.S., Kim, C.: Increasing camera pose estimation accuracy using multiple markers. In: Proc. of International Conference on Artificial Reality and Telexistence. (2006) 239–248
16. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification (2nd Edition). Wiley-Interscience (2000)

Using Photo-consistency for Intra-operative Registration in Augmented Reality based Surgical Navigation

G. Gonzalez-Garcia and R.J. Lapeer

School of Computing Sciences, University of East Anglia, UK

Abstract. We investigated the potential of an intensity-based metric, known as photo-consistency, for the intra-operative registration of virtual and real models in an Augmented Reality based Surgical Navigation system. We evaluated the accuracy of three different photo-consistency metrics described in the literature in combination with a number of optimisation methods including Powell's method, Differential Evolution and CODEQ. A lab scenario with a stereo pair of images of a phantom skull as seen through a stereoscopic surgical microscope and images from its corresponding CT-scanned virtual counterpart were pre-registered using an iterative closest point (ICP) algorithm. To simulate accumulating tracking errors, the virtual images were offset in x-y and the proposed photo-consistency method was then applied to re-register the real and virtual images. The root mean squared distance (RMS) to the original ICP registration (used as ground truth) was then calculated and its trend observed during optimisation of photo-consistency. Both metrics, that is photo-consistency and RMS showed on average a decreasing trend throughout the optimisation process, indicating that photo-consistency could be used as a metric for intra-operative registration in Augmented Reality based Surgical Navigation when no ground truth is available.

1 Introduction

Augmented Reality (AR) based or Image Enhanced Surgical Navigation (IESN) systems aim to provide the surgeon with enhanced visual information as compared to the standard view obtained by optical devices such as the surgical endoscope used in Minimally Invasive Surgery (MIS) and the stereoscopic surgical microscope. This is achieved by overlaying pre-operatively scanned imagery through X-ray Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) on top of the captured images. The resulting effect produces the so-called "X-ray" vision in which the practitioner can perceive hidden structures or organs not noticeable by the "naked" eye.

The procedure starts with *calibration* of the camera(s) that will capture the imagery upon which the virtual models will be superimposed. This is followed by a procedure known as *registration* which is required to align both the real images and CT/MRI models. Both procedures are performed pre-operatively and require some kind of calibration pattern attached to the patient. They eventually produce a static overlay between models. The third stage encompasses *tracking* of the position of the cameras and/or patient to reflect their movements during surgery. One of the problems found in IESN is that each independent stage produces a certain level of error in overlay accuracy. When these stages are combined, the resulting overlay accuracy is affected by the accumulated errors and is aggravated by the use of the tracking device over a prolonged time, which eventually results in an unacceptable overlay accuracy.

To alleviate this problem without having to resort back to the initial calibration/registration procedure and setup, we investigate a registration method which can be performed intra-operatively and is based solely on the visual information obtained from a pair of cameras connected to a surgical microscope. The registration technique involves a cost function that compares the difference in intensity values between the captured images and evaluates the matching accuracy through a metric denominated photo-consistency.

Photo-consistency is a constraint that has been used for the reconstruction of 3D models from a set of colour or greyscale images in which the real scene is subdivided into voxels [1]. Clarkson et al. [2] employed this method as a similarity measure to match the projection of two or more 2D images to a 3D surface model using calibrated cameras and an optimisation function based on gradient ascent search. Later, Jankó and Chetverikov [3] generalised the technique by finding the registration pose and performing a camera calibration procedure at the same time through a genetic algorithm. Both research groups employed full-sized polygonal surface models as target objects.

In the medical field, Chen et al. [4] implemented a photo-consistency cost function to intra-operatively register calibrated endoscopic images to volumetric models. Several images were captured by placing the endoscope at different positions and the optimisation was performed using Powell's method. A different registration approach was presented in [5] using a calibrated stereo endoscope in a video sequence of a beating heart. Although in this setup the endoscope position was fixed, the real model was under the influence of a heart cycle motion which required multiple CT reconstructions. Their method was used to register full-sized real and virtual models.

2 Method

2.1 Estimation of photo-consistency

The estimation of photo-consistency relies on the comparison between colour or intensity values in a set of an object's 3D points which are projected on two or more image viewpoints of the same scene. Therefore, the corresponding pixels related to a same point should ideally have the same colour/intensity properties on each image. If the difference between related pixel values is null or near zero, it is said that the point is photo-consistent in the set of views. One of the constraints of this method is that the visible object must maintain an equal luminance regardless of the point of view — a Lambertian model.

In our current setup, we connect two black and white cameras to the optics of a surgical microscope for Ear, Nose and Throat (ENT) procedures. Each camera is calibrated pre-operatively using Tsai's algorithm [6] which produces a projection matrix $P = KM$, relating the internal parameters (K) and the external parameters (M). Thus, the projection of a model's 3D point $\mathbf{X}$ on each camera image is calculated as: $x_l \sim P_l\mathbf{X}$ and $x_r \sim P_r\mathbf{X}$ where P_l and P_r are the 3 x 4 projection matrices for the left and right cameras, respectively; and x_l and x_r are the corresponding projected pixels of the same point $\mathbf{X}$. The sign $\sim$ signifies that the projection is defined up to a scale factor.

We determine the value of the photo-consistency function PC by comparing the pixel intensity levels in the pair of captured images:

$$PC = \frac{1}{N}\sum_{i=1}^{N} \|I(x_{l,i}) - I(x_{r,i})\|^2 \tag{1}$$

where N represents the total number of visible pixels i in both images. Clarkson et al. [2] provide an alternative similarity measure for the computation of photo-consistency by first determining a mean of pixel values. In the case of two viewpoints it follows as: $I(\bar{x}) = (I(x_{l,i}) + I(x_{r,i}))/2$. Then, the total sum of squared differences is computed according to the following equation:

$$PC_{sq} = \frac{1}{N}\sum_{i=1}^{N} \frac{\left(I(x_{l,i}) - I(\bar{x})\right)^2 + \left(I(x_{r,i}) - I(\bar{x})\right)^2}{2} \tag{2}$$

A final cost function also described in [2], aims to reduce the effect of outliers by calculating the inverse of squared differences. This is achieved by using a threshold ϵ related to the noise level found in intensity images:

$$PC_{inv} = \frac{1}{N}\sum_{i=1}^{N} \frac{\epsilon^2}{\epsilon^2 + \left(\left(I(x_{l,i}) - I(\bar{x})\right)^2 + \left(I(x_{r,i}) - I(\bar{x})\right)^2\right)} \tag{3}$$

2.2 Intra-operative registration

In order to align both real and virtual models at the beginning of the surgery it is necessary to perform an initial registration. In our case we use the Iterative Closest Point (ICP) algorithm [7] to match a VBH mouthpiece [8] worn by the patient during the pre-operative scan and at the time of surgery. However, due to accumulating errors throughout the medical procedure the original alignment tends to get worse as described earlier. For this purpose, we apply a photo-consistency based cost function to evaluate the best registration pose that corresponds to the lowest intensity difference between images. Because the cameras have been calibrated in an earlier stage, the cost function only requires determining the extrinsic parameters, i.e. three translational and three rotational degrees of freedom.

Firstly, a set of visible voxels in the virtual model is selected by backprojecting screen pixels of that model within a user-defined selection window. This technique is similar to a raycasting projection. A voxel is then detected for each projected ray which collides on the volumetric model - see Figure 1. As several screen pixels will map to one voxel due to magnification of the microscope, duplicated voxels are ignored. Secondly, a forward projection ray is created from the selected voxel to each of the cameras in order to determine the corresponding pixel coordinates in both 2D images.

During the projection from voxel to pixel coordinates it is necessary to evaluate any possible voxel occlusion that obstructs visibility of a point on both viewports. For surface mesh models a z-buffer approach can be implemented to render only the external visible points [2]. Surface normals can also be computed to avoid comparing areas of the surface that are not oriented towards the camera [3]. However, as CT or MRI models are made of voxels with different transparency levels, we have decided to use a direct check of possible voxels that can partially obstruct the forward-projected ray from a selected 3D point. Therefore, if the ray collides with a voxel that has a higher transparency level than a specific threshold, we determine it as an occlusion and the corresponding projected pixels are not taken into account for the evaluation of photo-consistency. As mentioned earlier, a projected voxel does not relate to a single pixel on the pair of captured images (voxel-to-pixel relation is 1:many) due to magnification. For this purpose it is required to determine the voxel dimensions on the model and project the vertices of the voxel face that is oriented towards each camera. From these four vertices it is possible to create a sub-window that relates to the visible voxel hence determining its corresponding pixels. We then apply a median filter to the pixels inside the convolution window.

2.3 Optimisation method

The optimisation method aims to minimise the cost functions described in Eqs 1-3 by iteratively changing the registration position for a number of iterations until a global minimum value is found. The optimised result must correspond to the best matching pose between the real and virtual anatomy. Several optimisation methods were tested which do not require the use of derivatives as the shape of the global function is unknown. We applied Powell's method and a number of genetic programming strategies based on Differential Evolution (DE) [9]. DE aims to optimise a problem by combining the individuals from a set of potential solutions. This combination selects the best existing candidates and arithmetically mixes them until the best value is found. However, its main drawback is tuning the initial control parameters for each different problem, e.g. in [10] the authors propose ten different strategies depending on the problem features. Also, a wrong selection of initial parameters can hinder the performance of the algorithm. Therefore, we chose a technique proposed by Salman et al. called Self-adaptive Differential Evolution (SDE) [11]. SDE eliminates most of the manual selection by constantly changing the control parameters and selecting the ones that produce the best results. This avoids stagnation in local minima and exploits a broader search in the function space. However, a minimum of two parameters are still needed to initialise the optimisation. Another approach we have investigated is a stochastic algorithm called CODEQ [12] which is based on

four different optimisation strategies: Chaotic search, Opposition-based Learning, Differential Evolution and Quantum mechanics. Its main advantage consists on being a parameter-free method that adapts itself to the objective function.
We implemented the three different optimisation methods, i.e. Powell, SDE and CODEQ in our image enhanced surgical navigation system software. We then tested the implementations on four benchmark functions as reported from the literature[1] which are used in the field to test optimisation algorithms. All three implementations passed this test which is crucial to their successful application and validation of the photo-consistency method.

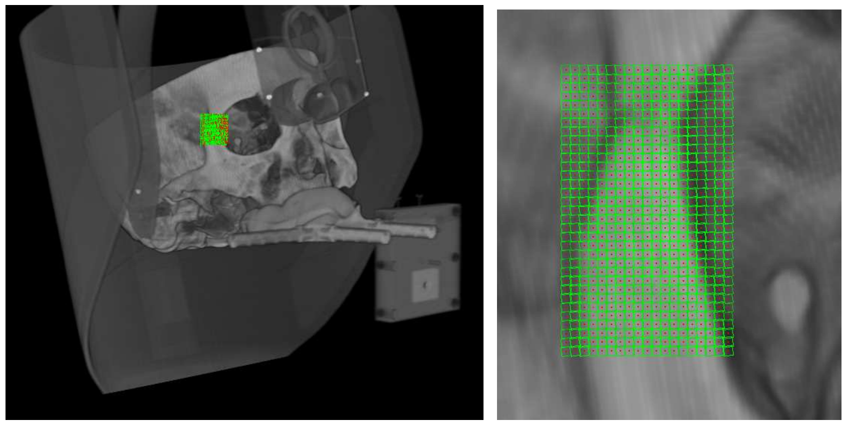
3 Experiment

The aim of the current experiment is to assess photo-consistency as an intra-operative registration metric to compensate for increasing registration errors due to tracking after initial ICP registration. In order to evaluate the intra-operative registration accuracy we used a human skull as a dummy patient. The procedure is largely as follows: After calibrating the cameras connected to the microscope and carrying out the initial ICP registration, we move our region of interest (ROI) towards one of the eye sockets (calibration and ICP registration are performed on the attached VBH mouthpiece). We then select a number of voxels on the model by back-projecting pixels in a selection window and record their 3D position - see Figure 1. This is followed by an initial intensity evaluation to calculate the photo-consistency in the current pose. The reason for the latter operation is because the initial photo-consistency value obtained does not necessarily give a zero value even though the real and virtual objects match closely. This is due to the error on the ICP registration (1-2mm) and the fact that the illumination produced by the microscope is not strictly Lambertian. For this reason, the 3D position and corresponding photo-consistency values at the aligned pose (after ICP registration) are used as ground truth to assess the registration results. Then, the CT-based virtual model is offset in x-y for the purpose of simulating a misregistration during surgery due to tracking. We compared the registration accuracy of different cost functions proposed in the literature for photo-consistency as discussed in section 2.1, i.e. Equations 1-3 and the three optimisation methods, i.e. SDE, CODEQ and Powell's method as discussed in section 2.3. The root mean squared distance (RMS) error was computed between the voxels located at the ground truth pose and at the final registration position for validation.

4 Discussion

Table 1 shows the optimal values of each combination of cost function and optimisation method. The best correlations between the RMS distance and PC metrics are found for the SDE and CODEQ optimisation methods, each combined with the PC and PC_{sq} cost functions. The PC_{inv} cost function decreases throughout optimisation but its lowest values do not correspond to the best RMS distances.

[1] The four selected benchmark functions are either unimodal or multimodal functions - we refer the reader to [11] for a description of these functions.



(a) (b)

Fig. 1. (a) selected voxels on virtual skull (CT); (b) zoom-in of selected voxels on virtual skull; prior to photo-consistency optimisation.

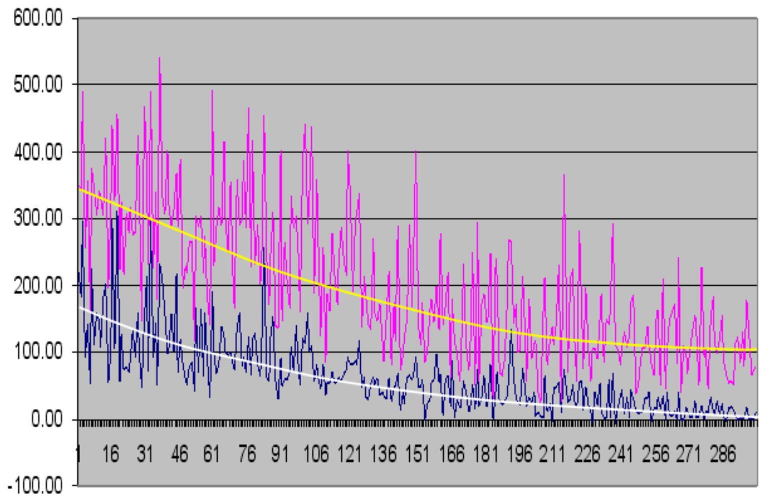


Fig. 2. Plot of the SDE-PC combination for ΔPC (magenta/top) and the RMS distance (blue/bottom; in mm scaled $\times 100$) as a function of number of iterations. The fitted smooth curves illustrate the correlation between both metrics and convergence.

	PC(ICP)	RMS(ICP)	Offset ΔPC	Offset RMS	Final ΔPC	Final RMS
SDE-PC	166.356	0.00	+100.411	2.83	+3.58725	0.23
SDE-PC_{sq}	6.16788	0.00	+1.87290	2.83	+0.196333	0.21
SDE-PC_{inv}	0.001239	0.00	+0.000492	2.83	+0.000011	1.18
CODEQ-PC	163.929	0.00	+101.500	2.83	+0.511383	0.17
CODEQ-PC_{sq}	6.43963	0.00	+1.85074	2.83	+0.004411	0.12
CODEQ-PC_{inv}	0.001239	0.00	-0.000646	2.83	+0.000255	1.4
Powell-PC	166.821	0.00	+100.543	2.83	-0.008820	1.20
Powell-PC_{sq}	6.40028	0.00	+1.77622	2.83	+0.001663	0.93
Powell-PC_{inv}	0.001156	0.00	-0.000564	2.83	+0.000027	6.68

Table 1. Photo-consistency (PC) and RMS values in mm. The first two columns correspond to the ground truth (ICP) registration. The next pair to the x-y translational offset to simulate accumulating tracking errors. The final pair shows the two metrics after optimisation using each of the nine combinations of PC based cost functions and optimisation methods.

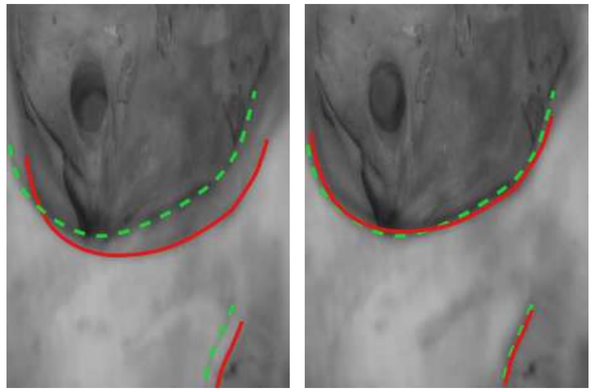
Powell's method generally fails to reach a global minimum which comes to no surprise considering its inability to explore the error surface. Occasionally, negative ΔPC's were found near the end of the optimisation process which indicate that better registrations than the original ICP were found. However, as we used the latter as the ground truth in this experiment, save for visual inspection, there is no qualitative evidence of this. Figure 2 shows the trend of the PC versus RMS metrics for the SDE-PC combination. The stochastic nature of the DE type optimisation algorithms is reflected in the spikiness of the curves, however on average they converge to a minimum (smooth curves). Fig. 3 shows the overlay before and after photo-consistency based registration.

5 Conclusion

We have illustrated the potential of photo-consistency as a cost function for intra-operative registration to compensate for tracking errors in x-y translation. Though perhaps even capable of yielding better results than the static ICP procedure, the current experiment is not conclusive on this assertion, due to the ICP registration being used as the ground truth. The SDE and CODEQ optimisation methods gave good accuracy but the former is substantially faster hence more suited for surgical navigation applications. Both PC and PC_{sq} metrics were shown to be useful cost functions, whereas the PC_{inv} metric did not reach convergence in most trials. Further experiments on the remaining four DOF's (rotation and z-translation) are currently in progress.

References

1. Seitz, S., Dyer, C.: Photorealistic scene reconstruction by voxel coloring. Int J Comput Vision **35**(2) (1999) 151 – 173
2. Clarkson, M., Rueckert, D., Hill, D., Hawkes, D.: Using photo-consistency to register 2D optical images of the human face to a 3D surface model. IEEE Transactions on Pattern Analysis and Machine Intelligence **23**(11) (2001) 1266 – 1281



(a) before registration (b) final registration

Fig. 3. Registration of real and virtual models around the skull orbit using PC_{sq} and CODEQ. Green/dashed lines show contour features in the real model. Red/solid lines indicate contour features in the virtual model.

3. Jankó, Z., Chetverikov, D.: Photo-consistency based registration of an uncalibrated image pair to a 3D surface model using genetic algorithm. In: Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization, and Transmission. (2004) 616 – 622
4. Chen, M., Gonzalez, G., Lapeer, R.: Intra-operative registration for image enhanced endoscopic sinus surgery using photo-consistency. MMVR15 proceedings, Studies in Health Technology and Informatics **125** (2007) 67 – 72
5. Figl, M., Rueckert, D., Hawkes, D., Casula, R., Hu, M., Pedro, O., Zhang, D., Penney, G., Bello, F., Edwards, P.: Registration of a 4D cardiac motion model to endoscopic video for augmented reality image guidance of robotic coronary artery bypass. AMI-ARCS workshop of MICCAI 2008 (September 2008)
6. Tsai, R.: A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. IEEE J. Robotics and Automation **3**(4) (1987) 323 – 344
7. Besl, P., McKay, N.: A method for registration of 3D shapes. IEEE Trans. Pattern Analysis and Machine Intelligences **14**(2) (1992) 239 – 256
8. Martin, A., Bale, R., Vogele, M., Gunkel, A., Thumfart, W., Freysinger, W.: Vogele-Bale-Hohner mouthpiece: Registration device for frameless stereotactic surgery. Radiology **208** (1998) 261–265
9. Storn, R., Price, K.: Differential evolution - A simple and efficient adaptive scheme for global optimization over continuous spaces. Technical Report TR-95-012, International Computer Science Institute (1995)
10. Price, K., Storn, R., Lampinen, J.: Differential Evolution - A Practical Approach to Global Optimization. Springer, Berlin-Heidelberg (2005)
11. Salman, A., Engelbrecht, A., Omran, M.: Empirical analysis of self-adaptive differential evolution. Eur J Oper Res **183**(2) (2007) 785 – 804
12. Omran, M., Salman, A.: Constrained optimization using CODEQ. Chaos, Solitons and Fractals **42** (2009) 662–668

EKF Monocular SLAM 3D Modeling, Measuring and Augmented Reality from Endoscope Image Sequences

Oscar G. Grasa[1], J. Civera[1], A. Güemes[2], V. Muñoz[3], and J.M.M. Montiel[1]

[1] I3A. Universidad de Zaragoza, Spain {oscgg,josemari}@unizar.es,
[2] Hospital Clínico Universitario "Lozano Blesa", Universidad de Zaragoza.
[3] Dpto. Informática e Ingeniería de Sistemas, Universidad de Malaga, Spain. *

Abstract. In recent years monocular SLAM has produced algorithms for robust real-time 3D scene modeling and camera motion estimation which have been validated experimentally using low cost hand-held cameras and standard laptops. Our contribution is to extend monocular SLAM methods to deal with images coming from a hand-held standard monocular endoscope. With the endoscope image sequence as the only input to the algorithm, a sparse abdominal cavity 3D model –a 3D map– and the endoscope motion are computed in real-time.
A second contribution is to exploit the recovered sparse 3D map and the endoscope motion to: 1) produce real-time photorealistic 3D models that ease cavity visualization; 2) measure distances in 3D between two points of the cavity; and 3) support augmented reality (AR) annotations. All this information can provide useful support for surgery and diagnose based on endoscope sequences. The results are validated with real hand-held endoscope sequences of the abdominal cavity.

1 Introduction

SLAM (Simultaneous Localization and Mapping) is a classical problem in mobile robotics: let be a mobile sensor following an unknown trajectory in an unknown environment, the goal is to estimate, simultaneously, both the environment structure –a map of 3D points– and the sensor location with respect to that map. Recently, SLAM research has focused on monocular cameras as the unique sensorial input, giving origin to monocular SLAM methods. 30 Hz real-time systems estimating full 3D camera motions and maps of 3D points using commodity cameras and computers have been reported in [1–4].

Our contribution is to extend, and validate with real monocular endoscope image sequences, one of the leading edge monocular SLAM algorithms [4] for its

* The authors would like to thank Dr. M. A. Bielsa form the Hospital Clínico Universitario "Lozano Blesa" for the "Abdominal wall image sequence". We would also like to thank Jonathan Richard Shewchuk for allowing free use of his Delaunay's Triangulation software (http://www.cs.cmu.edu/~quake/triangle.html). This work has been partially supported by Spanish FIT-360005-2007-9, DPI2006-13578 and DPI2009-07130, and EU funded RAWSEEDS Grant FP6-IST-045144.

use in medical applications. The primary result is a sparse 3D map composed of salient points –features– of the observed cavity. The main assumptions are that the cavity is rigid and that the endoscope undergoes a non-pure rotational motion. These conditions are fulfilled by a number of medical applications, such as laparoscopic ventral hernia repairs. The proposed work expands and details the abstract recently published by the same authors, [5].

Monocular SLAM methods recover the map up to an unknown scale factor, implying that only relative distances can be measured. However, in practice, a known size tool can provide the unknown scale factor and hence real distances can be recovered. Given the probabilistic nature of the SLAM map, the distance estimates are accompanied by an error estimate. We show how relative distances, along with the corresponding error estimates, are computed in live real-time while exploring a cavity.

Besides, the map is used as the backbone for real-time photorealistic modeling to ease the 3D cavity visualization. The textured 3D model allows the synthesis of a panorama that expands the limited FOV (field of view) of the endoscope. Finally, since the camera motion with respect to the 3D map is known accurately and in real-time, AR annotations can be supported live in medical sequences.

A number of techniques have been developed for cavity 3D reconstruction from endoscope sequences. [6,7] perform reconstructions using stereo endoscopes. In both works, the stereo endoscope always points to the same cavity area or moving organ to obtain the 3D structure. Stereo endoscopes have been successfully used in visual SLAM as reported in [8], where the performance of different image features when medical images are considered is analyzed. This is the closest work to ours. However, we are able to deal with monocular images.

Regarding the usage of monocular computer vision to determine 3D models from endoscope sequences, a significant paper is [9], where cavity 3D structure is computed to align it with a cavity CT scan model. More recently, in [10], classical two view RANSAC+bundle adjustment is applied to mannequin images to determine the 3D structure. This work presents a constraint-based factorization 3D modeling method from endoscope sequences to produce a dense 3D reconstruction in near real-time. Compared with these works, we demonstrate experimentally that our monocular SLAM algorithm features robust real-time performance when processing real endoscope images.

Monocular SLAM has proven a right tool for providing scene anchor points and camera motion estimation at frame rate in [2,11]. In this work we show that AR can be supported when monocular SLAM is adapted to medical images.

2 EKF Monocular SLAM

We focus on the EKF+ID+JCBB monocular SLAM approach. EKF (Extended Kalman Filter) monocular SLAM was initially proposed in [1]. ID (Inverse Depth) for map point coding, [4], improves measurement equation linearity and hence the overall estimation performance. JCBB (Joint Compatibility Branch and Bound) [12] has proven essential for spurious rejection by enforcing scene rigidity. This combination was first used for monocular cameras in [13].

The estimated state is a Gaussian vector $\mathbf{x}$ that jointly codes the camera location with respect to the map, $\mathbf{x_v}$, and all map points coded in inverse depth, $\mathbf{y_i}$. The camera smooth motion prior is coded by means of a constant velocity motion model. Because of that, the camera location $\mathbf{x_v}$ includes: translation ($\mathbf{r^W}$), orientation defined by a quaternion ($\mathbf{q^{RW}}$), velocity ($\mathbf{v^W}$), and angular velocity ($\omega^{\mathbf{R}}$). The Gaussian estimate is defined by its mean, $\hat{\mathbf{x}} = \left(\hat{\mathbf{x}}_\mathbf{v}^\top \; \hat{\mathbf{y}}_\mathbf{1}^\top \; \hat{\mathbf{y}}_\mathbf{2}^\top \; \ldots \right)^\top$, and covariance, P. Monocular sequence processing can recover 3D structure up to an unknown scale factor and up to an unknown 3D transformation with respect to an absolute reference. To remove the absolute transformation from the estimate, the first camera is defined as the absolute reference W (see Fig. 1).

2.1 Inverse Depth

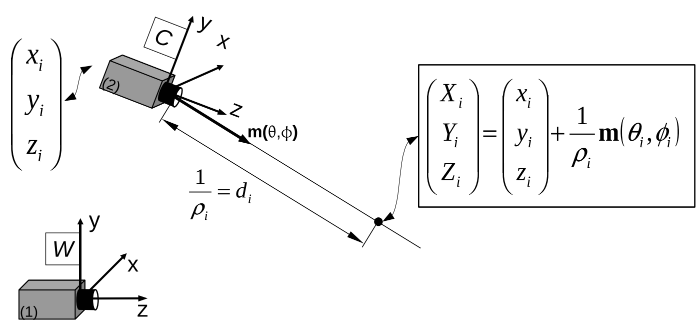


Fig. 1. Camera-(1) defines the world frame, W. A feature is observed for the first time by camera-(2), the feature location is defined with respect to the camera-(2) pose, $(x_i, y_i, z_i)^T$, using distance between camera-(2) and the feature, $d_i = 1/\rho_i$, and a directional vector, $\mathbf{m}(\theta, \phi)$, defined by its azimuth and elevation angles.

ID point coding [4] improves the measurement linearity –and hence the EKF performance– at low parallax, which takes place when feature depth is much bigger than the camera translation. At initialization, even features close to the camera produce low parallax. As a result, ID improves performance even for maps composed of close features only. An ID feature is a 6 parameters vector:

$$\mathbf{y_i} = \left(x_i \; y_i \; z_i \; \theta_i \; \phi_i \; \rho_i\right)^T \tag{1}$$

The projection ray of a map point when it is first observed is coded as: x_i, y_i, z_i (camera location when the point was observed for the first time), and θ_i and ϕ_i (azimuth and elevation angles), which define the ray unit vector $\mathbf{m}(\theta_\mathbf{i}, \phi_\mathbf{i})$. Point depth is coded by its inverse $\rho_i = 1/d_i$, so a point location $\mathbf{x_i}$ is (see Fig.1):

$$\mathbf{x_i} = \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} = \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} + \frac{1}{\rho_i}\mathbf{m}(\theta_i, \phi_i), \quad \mathbf{m}(\theta_i, \phi_i) = \begin{pmatrix} \cos\phi_i \sin\theta_i \\ -\sin\phi_i \\ \cos\phi_i \cos\theta_i \end{pmatrix} \tag{2}$$

2.2 Data Association and Feature Detection

The quality of the SLAM reconstruction strongly depends on the data association accuracy. Active matching in monocular SLAM combines an accurate geometrical prediction with an image correlation score. EKF innovation defines for every map feature a predicted location and an elliptical uncertainty region. The match is searched for inside this area by correlation with a texture patch, which in our case is 11x11 pixels in size, stored when the map point is first observed and warped according to the point-camera relative location. As all innovations are highly correlated through the camera location uncertainty, JCBB is applied, checking if all matches are jointly compatible. If not, a Branch and Bound algorithm is applied to detect inconsistent matches before the EKF update.

Monocular SLAM in robotics uses a correlation score based on luminance, neglecting color information. However, internal organ images have high red and low blue content, so we use the green band, which contains a rich contrast to fire a point detector and a nice texture to produce distinctive patches for recognition.

To initialize a map point, a Harris saliency detector is applied. In endoscope scenes the light source is fixed to the camera producing reflections that erroneously fire the detector. To remove these reflections we assume they produce high gray level pixels, and if any pixel in a patch around the feature is over a threshold, the point is rejected. We use a theshold of 140 over 255.

3 SLAM Geometrical Map Exploitation

Monocular SLAM produces a sparse 3D map and the camera motion along with the corresponding covariance. We propose using the SLAM map as the geometrical backbone to support useful information for medical applications.

3.1 Photorealistic reconstruction

A mesh of triangular elastic textured tiles is built on top of the SLAM map. It is a generalization for 3D scenes of the mosaic method proposed in [14]. The tiles are defined by a standard 2D Delaunay's triangulation over a projection of the 3D map on the XY plane of the absolute reference W.3D triangle texture is gathered from the images that observe the complete corresponding 3D triangle. Fig. 2 sketches the photorealistic modeling process.

Since triangulation is a live process –map points, and consequently triangles, are continuously created, erased and their estimates changed–, maintenance operations are performed to deal with new and deleted triangles as the SLAM estimation evolves, and to take textures for the triangles from the images.

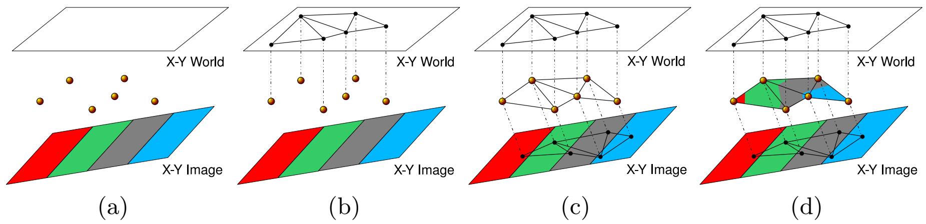


Fig. 2. *a*: Features in 3D, the current image and the world X-Y plane. *b*: Features projected onto a X-Y plane and triangulation on this plane. *c*: Triangulation image backprojection to obtain textures. *d*: Final photorealistic reconstruction.

3.2 Distance Measurement

Distance measurements are needed for some surgical procedures, e.g. ventral hernia repairs where the hole size is used to define the prothesic mesh dimensions.

Given the probabilistic 3D map, up to a scale factor, and an element of known size, for example a tool, real distances between points of the map, along with measurement error estimates, are available.

Considering two reference points, (r_1, r_2), at a known distance, which define the scale factor, s, the distance between two map points (i, j) is:

$$d(i,j) = s\frac{d_m(i,j)}{d_m(r_1,r_2)} \tag{3}$$

where $d_m(i,j)$ and $d_m(r_1,r_2)$ are the Euclidean distances between points (i,j) and reference points (r_1, r_2), respectively, measured in the SLAM map.

As the distance is a function of the SLAM state vector, $\mathbf{x}$, the covariance of the distance estimation can be propagated linearly from the SLAM covariance by means of the corresponding Jacobian matrix, $\mathbf{J}$:

$$\mathbf{J} = \frac{\partial d(i,j)}{\partial \mathbf{x}}, \mathbf{x} = \left(\mathbf{x_v}^\top\ \mathbf{y_1}^\top \ldots \mathbf{y_{r1}}^\top \ldots \mathbf{y_{r2}}^\top \ldots \mathbf{y_i}^\top \ldots \mathbf{y_j}^\top \ldots\right)^\top \tag{4}$$

where all features are coded in ID. Since $d(i,j)$ only depends on i, j, r_1 and r_2, $\mathbf{J}$ is sparse, and reduced Jacobian $(\mathbf{J}_r)$ and covariance $(\mathbf{P_r})$ matrices are used instead of the full matrices to compute the measurement error estimate σ_d^2:

$$\sigma_d^2 = \mathbf{J_r}\mathbf{P_r}\mathbf{J_r}^\top, \quad \mathbf{P_r} = \begin{pmatrix} P_{y_{r1}y_{r1}} & P_{y_{r1}y_{r2}} & P_{y_{r1}y_i} & P_{y_{r1}y_j} \\ P_{y_{r2}y_{r1}} & P_{y_{r2}y_{r2}} & P_{y_{r2}y_i} & P_{y_{r2}y_j} \\ P_{y_iy_{r1}} & P_{y_iy_{r2}} & P_{y_iy_i} & P_{y_iy_j} \\ P_{y_jy_{r1}} & P_{y_jy_{r2}} & P_{y_jy_i} & P_{y_jy_j} \end{pmatrix} \tag{5}$$

3.3 Augmented Reality

AR annotations in endoscope images need accurate real-time estimates of the live camera motion with respect to the observed scene. Monocular SLAM based only on images gathered by a camera has proven capable of providing camera motion in real-time at 30 Hz, [2,11], for rigid scenes. AR is useful in laparoscopic surgery because it enables to visualize notations and fuse other modal images, such as 3D models of CT or MR, with endoscope images live during surgery.

Our contribution is to show that EKF monocular SLAM can be successfully applied to support AR annotations using as only input image sequences corresponding to a real hand-held endoscope observing the abdominal cavity.

4 Results

Experimental validation is performed over real images (360×288) at 25 Hz gathered by a hand-held monocular endoscope observing the abdominal cavity, the image sequence being the only data input to the algorithm. Real-time performance at frame rate is achieved in all the experiments. Endoscope intrinsic parameters have been calibrated using a standard planar pattern calibration method, based on Zhang's initial solution [15], followed by Bundle Adjustment. A two parameter radial distortion model has been applied.

In Figs. 3 and 4 the textured triangular mesh model is shown. Despite the sparse map being composed of a reduced number of points, the photorealistic model provides easy understanding of the 3D cavity structure. Fig. 4 shows another example of a 3D live photorealistic model estimate corresponding to the abdominal wall, during a hernia repair surgery. The video "reconstruction.avi" shows the live model estimation and a detailed visualization of the 3D model from different points of view to assess the quality of the model.

Distances, together with their errors, have been calculated relative to a scale. The conversion to real distances is immediate if the real distance between two points in the reconstruction is known, as the scale can then be resolved (see Fig. 3). Fig. 3(b) shows the estimation history both for the distance and the error. Initially, error uncertainty is big, but as the camera translates, point location error decreases and consequently the distance error decreases as well. As the uncertainty is computed in live real-time, visual feedback gives the surgeon information on how to move the camera in order to reduce the distance error (see video "measurement.avi").

Since the 3D map and the camera location with respect to the map are available in live real-time, it is possible to anchor AR annotations to map points. Fig. 4 shows an AR cylinder, both in 3D and superimposed on the endoscope live image. As the virtual insertions are fixed to the map, they can be observed at their real location even when they are out of the camera FOV. The video "augmentedReality.avi" shows the corresponding movie.

5 Discussion and Future Work

We have shown how cavity exploration with a hand-held monocular endoscope can be casted as a monocular SLAM problem. A sparse map of 3D features and the camera motion are computed in live real-time at 25 Hz using the endoscope

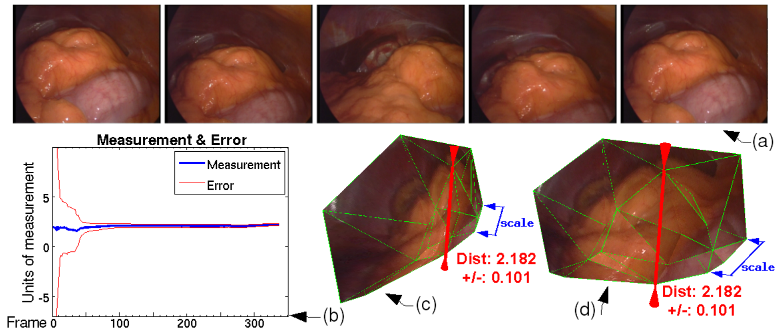


Fig. 3. Hand-held endoscope 341 frames sequence. *(a)* Several frames. *(c,d)* Photorealistic 3D model and distance estimate with the 2σ, 95%, error interval. *(b)* Historical distance and error estimates. Notice the error reduction as the camera moves and gathers information from different points of view providing higher parallax. See also "measurement.avi" in the additional material.

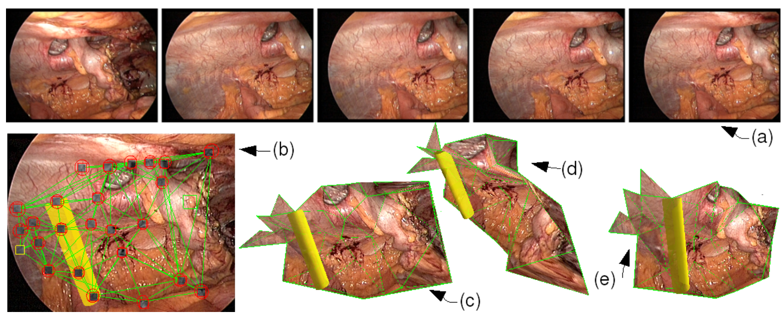


Fig. 4. Hand-held endoscope abdominal wall sequence. *(a)* Several frames. *(c,d,e)* Photorealistic 3D model and an augmented reality cylindrical insertion. *(b)* AR cylinder backprojected in live endoscope video. See also "reconstruction.avi" and "augmentedReality.avi" in the additional material.

image sequence as only input. The proposed algorithm is the state-of-the-art EKF+ID+JCBB monocular SLAM approach adapted to medical images.

It has also been shown how, building on the sparse SLAM map, a photorealistic model of the cavity can be computed in real time. The sparse 3D map provides support for AR annotations and 3D distance measurement. Our main contribution is to show this capabilities on real images gathered by a monocular endoscope observing the abdominal cavity.

Once we have tested the feasibility of the basic technique several venues of future work are open. As near future work, experiments to validate accuracy with respect to a ground truth in medical imagery would be valuable. Additionally, a cross-fertilization between engineers and physicians is needed to identify medical procedures that can benefit from the current state of the monocular SLAM technology; up to now we have focused on the abdominal cavity.

Current algorithms assume: 1) scene rigidity, 2) smooth endoscope motion, and 3) low motion clutter. These assumptions do not hold in general medical scenes: non rigidity is almost prevalent, sudden motions are frequent, and tools cause a significant motion clutter. Thus, future work is being directed to cope with these issues. Relocation algorithms such as [16, 17] can provide robustness with respect to non smooth motion or motion clutter. Further research is also needed in models that code the scene as a deformable one.

References

1. Davison, A.J.: Real-time simultaneous localisation and mapping with a single camera. In: International Conference on Computer Vision. (2003)
2. Klein, G., Murray, D.: Parallel tracking and mapping for small ar workspaces. In: Symposium on Mixed and Augmented Reality (ISMAR). (2007)
3. Eade, E., Drummond, T.: Unified loop closing and recovery for real time monocular SLAM. In: British Machine Vision Conference. (2008)
4. Civera, J., Davison, A., Montiel, J.M.M.: Inverse depth parametrization for monocular SLAM. IEEE Trans. on Robotics and Automation **24**(5) (2008) 932–945
5. García, O., Civera, J., Güemes, A., Muñoz, V., J.M.M., M.: Real-time 3d modeling from endoscope image sequences. In: Workshop on Advanced Sensing and Sensor Integration in Medical Robotics (ICRA2009)
6. Stoyanov, D., Darzi, A., Yang, G.Z.: A practical approach towards accurate dense 3d depth recovery for robotic laparoscopic surgery. Computer Aided Surgery (2005) 199–208
7. Mourgues, F., Devernay, F., Coste-Manière, É.: 3D reconstruction of the operating field for image overlay in 3D endoscopic surgery. In: IEEE/ACM Symp. Augmented. Reality. (2001) 191–192
8. Mountney, P., Stoyanov, D., Davison, A., Yang, G.Z.: Simultaneous stereoscope localization and soft-tissue mapping for minimal invasive surgery. In: Medical Image Computing and Computer-Assisted Intervention. (2006) 347–354
9. Burschka, D., Li, M., Taylor, R., D., H.G., Ishii, M.: Scale-invariant registration of monocular endoscopic images to ct-scans for sinus surgery. Medical Image Analysis **9**(5) (2005) 413–426
10. Wu, C., Sun, Y., Chang, C.: Three-dimensional modeling from endoscopic video using geometric constraints via feature poisitioning. IEEE Trans. on Biomedical engineering **54**(7) (2007)
11. Davison, A., Reid, I., Molton, N., Stasse, O.: Monoslam: Real-time single camera slam. IEEE Trans. on Pattern Analysis and Machine Intelligence **29**(6) (2007)
12. Neira, J., Tardos, J.D.: Data association in stochastic mapping using the joint compatibility test. IEEE Trans. on Robotics and Automation (2001) 890–897
13. Clemente, L.A., Davison, A.J., Reid, I.D., Neira, J., Tardós, J.D.: Mapping large loops with a single hand-held camera. In: Robotics Science and Systems. (2007)
14. Civera, J., Davison, A.J., Magallón, J.A., Montiel, J.M.M.: Drift-free real-time sequential mosaicing. Int. Journal of Computer Vision **51**(2) (2009) 128–137
15. Zhang, Z.: A flexible new technique for camera calibration. IEEE Trans. on Pattern Analysis and Machine Intelligence **22**(11) (2000) 1330–1334
16. Williams, B., Klein, G., Reid, I.: Real-time SLAM relocalisation. In: Proc. International Conference on Computer Vision. (2007)
17. Cummins, M., Newman, P.: FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance. The International Journal of Robotics Research **27**(6) (2008) 647–665

Demonstration Session – Summaries

Demonstration at AMI-ARCS 2009, London
Medical Contextual In-Situ Visualization

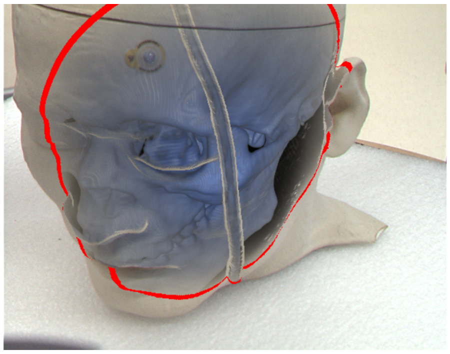
Christoph Bichlmeier[1], Maksym Kipot[1], Stuart Holdstock[1], Sandro Michael Heining[2], Ekkehard Euler[2], and Nassir Navab[1]

[1] Department of Chair for Computer Aided Medical Procedures & Augmented Reality, TU Munich, Germany
[2] Trauma Surgery Department, Klinikum Innenstadt, LMU, Munich, Germany

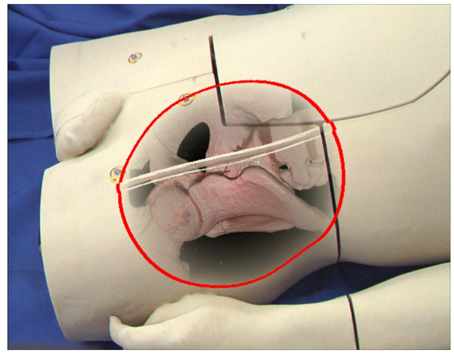
1 Introduction

Providing a reasonable composition of real and virtual entities within AR environments is a crucial task and a hot subject of recently published research. The objective of contextual in-situ visualization is to embed a virtual object in an Augmented Reality scene without loosing its spatial context to other objects of the scene. Position, shape, structure and the spatial order of involved objects have to be intuitively perceived without imposing additional mental workload. In particular, the composition becomes important when virtual objects are physically positioned behind real ones. This kind of object topology is given when medical imaging data such as CT or MRI is visualized in-situ, i.e. registered with the patient. However, if virtual anatomy is simply superimposed onto the patient, it appears to be floating above the real body. This results in misleading depth perception of entities of the AR scene and negatively affects the usability and acceptance of the system. The Chair for Computer Aided Medical Procedures & Augmented Reality, TU Munich, Germany is going to present a video based and a data based approach of contextual in-situ visualization. Both approaches create visual artificial depth cues, which improve the perceptual quality for both, passive observation of and interaction with the AR scene.

The data based approach uses primarily the geometry of the patient's skin to generate context information for better perception of deeper seated anatomy (see Fig. 2). However, for many interventional cases this is not practical and flexible enough. The majority of image guided surgeries does not use imaging data that is large enough to cover the skin geometry. Even if the skin geometry is available, here we deal only with only static context information captured at a certain point in time. For this reason, the skin serving as the major context layer cannot adapt to geometry deformation or color changes. The video based approach uses exclusively information from video data of video see-through devices to generate context information and introduce additional visual cues reacting to changing light and color conditions, deformation, and interaction with surgical instruments (see Fig. 2).

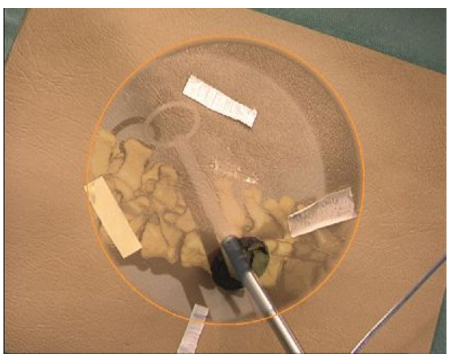


(a) Augmented head from CT data.

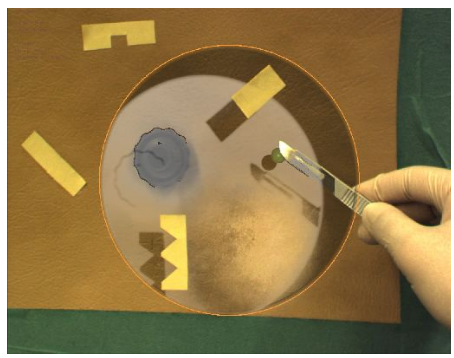


(b) Augmented hip and spinal column from CT data.

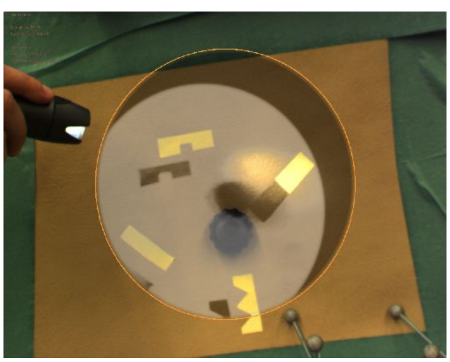
Fig. 1. The Visible Korean Human Phantom (VKHP) has been used as a testbed for the development of a data based approach of contextual in-situ visualization.



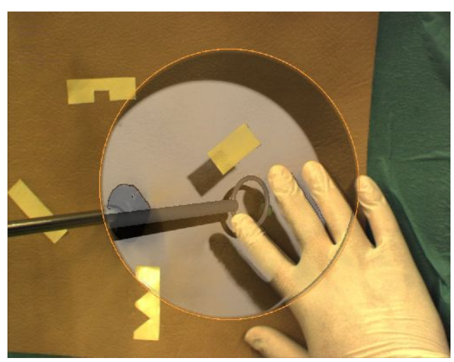
(a) Deformation of the skin



(b) Extracted instruments cast shadow



(c) Flashlight causes context information.



(d) Hand causes shadow feedback.

Fig. 2. Different features of the proposed video based contextual in-situ visualization.

Medical Augmented Reality using Autostereoscopic Image for Minimally Invasive Surgery

Hongen Liao

Graduate School of Engineering, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
liao@bmpe.t.u-tokyo.ac.jp

Abstract. One interesting field of imaging technology is naked-eye stereoscopy, which displays 3D autostereoscopic images without the need for special eyeglasses or tracking devices. A novel medical autostereoscopic image technique named Integral Videography (IV) will be introduced in this demonstration session. Because IV projects a 3D image into space, it has advantages over the traditional stereoscopic method, where different images are displayed for the viewer's left and right eyes. Using IV, the 3D image can be observed from a wide area in front of the display by several viewers at once. The development of relative image overlay techniques makes it appear that the 3-D image is inside the patient's body, and enables a medical augmented reality environment for minimally invasive surgery.

1 Introduction

Rapid technical advances in medical imaging, including its growing application in therapy and invasive/interventional procedures, have attracted significant interest in the close integration of research in the life sciences, medicine, physical sciences, and engineering. The obtained images are used to accurately identify treatment areas by acquiring pre-/intra-operative information and updating it to a navigation system used in iamge-guided therapy. Although current AR systems can adequately handle depth cues based on geometry (for instance, relative size, motion parallax, and stereo disparity), incorrect visualization of interposition between real and virtual objects has already been identified as a serious issue. We have developed an autostereoscopic imaging technique, in contrast to a 2D display or binocular stereoscopic display, which can be integrated into a surgical navigation system by superimposing an actual 3D image onto the patient. The autostereoscopic images are created by using a modified version of integral videography (IV) (Fig.1) [1], which reproduces 3D images using a micro convex lens array and a high-resolution high-pixel-density flat display. Using IV, the 3D image can be observed from a wide area in front of the display by several viewers at once (Fig.2). The development of relative image overlay techniques makes it appear that the 3-D image is inside the patient's body, and enables a medical augmented reality environment for minimally invasive surgery (Fig.3). This system could revolutionize real-time, autostereoscopic, in vivo imaging. Furthermore, this intriguing technology not only has applications in medicine, but is being studied as a practical technology for a variety of professional applications.

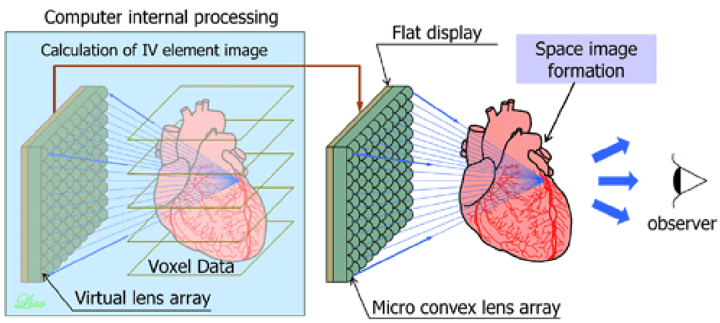


Fig. 1 Principle of integral videography: showing how to generate and reproduce a 3-D object by IV. Light rays reflected at the first point seen from the observer's side on the 3-D object pass through the centers of all lenses in the array and are redisplayed on the flat display. When the image is lit from behind, light rays intersect at the original point to become a new light point.

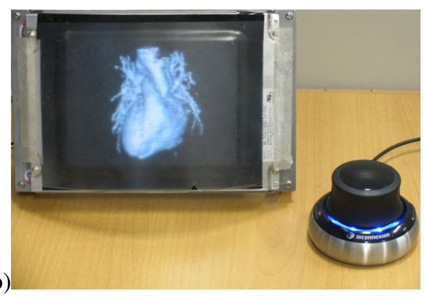
(a)  (b) 

Fig.2 (a) Example of autostereoscopic IV imaging device. (b) IV display system using high-speed rendering.

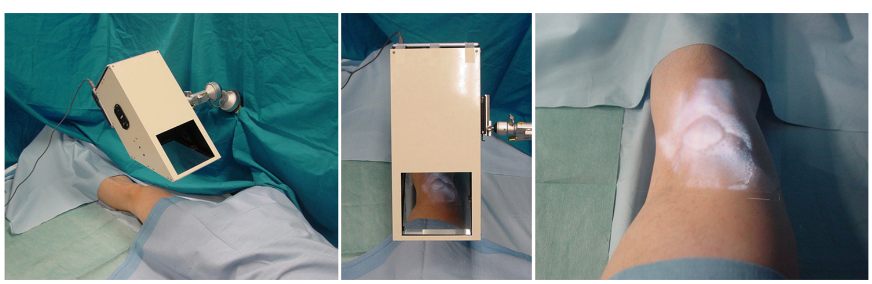


Fig.3 Example of autostereoscopic IV image overlay for orthopedics surgery. The photos show IV image overlay device and target with IV image overlay.

[1] H. Liao, et al: Surgical Navigation by Autostereoscopic Image Overlay of Integral Videography, IEEE T INF TECHNOL B., Vol.8 No.2, pp.114-121, 2004.

Miniaturised Augmented Reality System (MARS)

Philip "Eddie" Edwards[1], Aakif Tanveer[1], Tim Carter[2],
Dean Barratt[2], Dave Hawkes[2]

1. Department of Computing, Imperial College, London, UK
2. Centre for Medical Image Computing, University College, London, UK

History

The so-called miniaturised augmented reality system (MARS) was developed at the then computational imaging sciences group at King's College, London – the former home of the CMIC group. The initial system was constructed from two 15" monitors and beam-splitters. By calibrating the fixed position of the virtual screens and estimating the position of the viewer's eye, we can obtain an accurate stereo calibration.

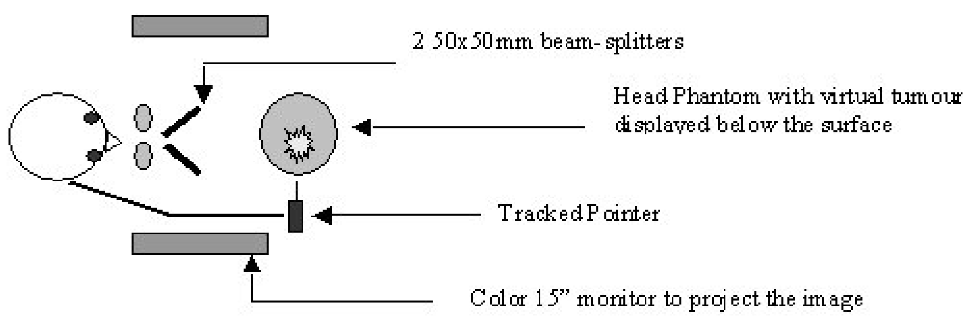
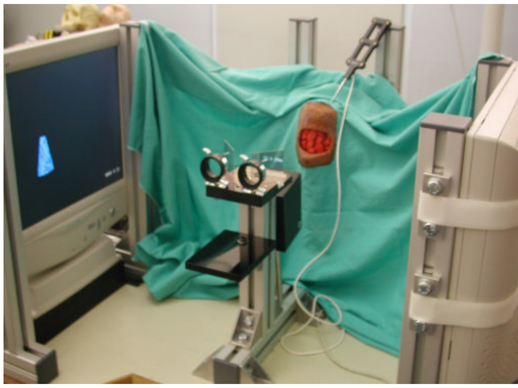


Figure 1. Diagram of the MARS

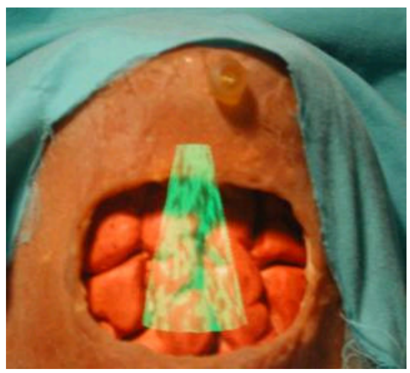
MARS

MARS is an optical see-through AR system built from off-the shelf components. Two 15" Philips flat screen LCD monitors (1024x768 resolution) provided the display devices. Equidistant from both monitor surfaces was a central block with two 50x50mm beam splitters, a pair of binocular eyepieces with all lenses removed and a chin rest to stabilise the observer's head and position the eyes in front of the eyepieces. The image from each screen projected to the beam splitter, where it was reflected through 90 degrees through the eyepiece to the observer's eye.

The MARS contained no lenses or magnifying optics so the image focal plane was manually set to be in the centre of the working display volume. MARS components were rigidly mounted to a bench on a portable trolley.



This system was intended as a test-bed for perception in stereo AR. Though MARS is anything but miniature, it provides one of the most compelling 3D AR experiences.



The intention in demonstrating this system at AMI-ARCS is to stimulate discussion about the quality of stereo AR that we should be aiming for and for attendees to be able to experience the perception that can be achieved from a high-contrast, high resolution, high brightness, accurately calibrated 3D AR system.

1. L. G. Johnson, P. J. Edwards, and D. J. Hawkes. Surface transparency makes stereo overlays unpredictable: The implications for augmented reality. In *Medicine Meets Virtual Reality*, Health Technology and Informatics, pages 131–136. IOS Press, 2003.

2. 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 *Proceedings of Medical Imaging and Augmented Reality*, Lecture Notes in Computer Science (3150), pages 369–376. Springer-Verlag, 2004.

ARView - Augmented Reality based Surgical Navigation system software for minimally invasive surgery

R.J. Lapeer and G. Gonzalez-Garcia

School of Computing Sciences, University of East Anglia, Norwich, UK

Description

ARView is an in-house developed software which provides an interface to a series of hardware devices typically used in surgical navigation for minimally invasive surgery, such as tracking devices and surgical visualisation aids (endoscopes, microscopes). It is based on our original 3D visualisation software, 3DView [1] - see Figure 1 - which is developed in C++/OpenGL (currently a DirectX version is underway). Additional to 3DView, ARView has built-in procedures for the three main operations needed to overlay real (intra-operatively acquired) and virtual (pre-operatively acquired) images: calibration of the optical system, registration of real and virtual imagery and tracking of any tools or objects/subjects which move in the operating scenery. The system has been proof-of-concept tested on a rigid sinus endoscope [2] and on a stereoscopic surgical microscope. Figure 2 shows a series of sinus endoscope images from a human skull model after calibration, registration and motion tracking operations for augmented reality based surgical navigation. Currently we are developing a new tracking system and are improving our registration procedure as registration and tracking operations are the main source of overlay error between real and virtual imagery.

References

1. 3DView: http://www.uea.ac.uk/cmp/research/medicalcomp/3dviewsoftware
2. Lapeer, R., Chen, M., Gonzalez, G., Linney, A., Alusi, G.: Image-enhanced surgical navigation for endoscopic sinus surgery: Evaluating calibration, registration and tracking. IJMRCAS **4** (2008) 32 – 45

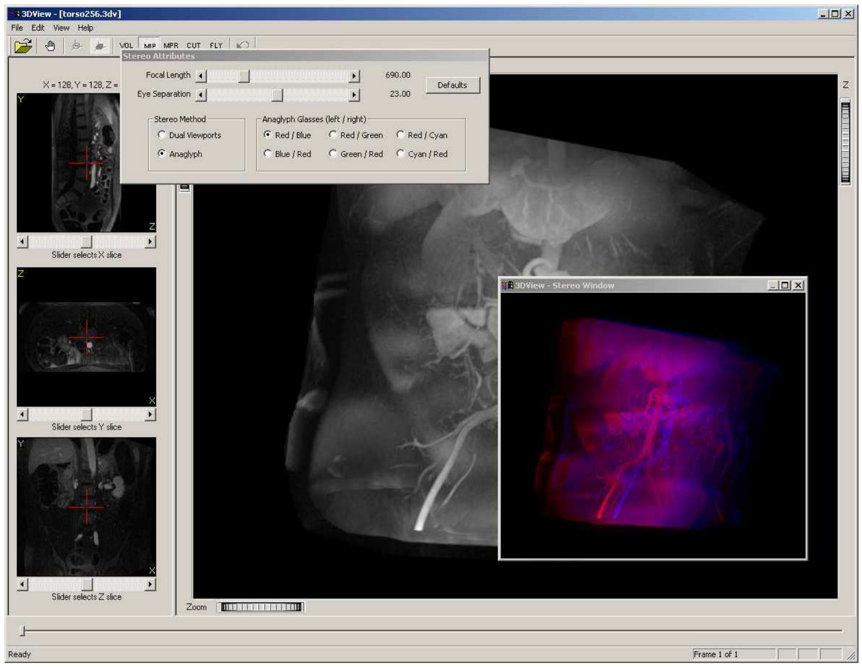


Fig. 1. A snapshot of the 3DView visualisation software.

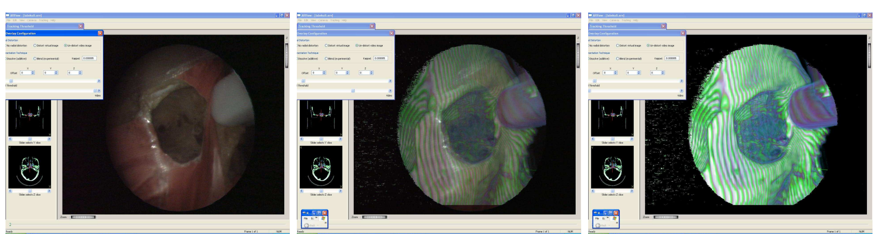


Fig. 2. Sequence of endoscope images after calibration, registration and motion tracking: left - real image (human skull orbit); middle - blended image; right - virtual image (CT scan of skull).

Perk Station – Percutaneous Surgery Training and Performance Measurement Platform

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

1. Queen's University, Kingston, ON, Canada
2. The Johns Hopkins University, Baltimore, MD, U.S.A
3. Worcester Polytechnic Institute, Worchester, MA, U.S.A

1. Introduction

Image-guided percutaneous (through the skin) needle-based surgery has become part of routine clinical practice in performing procedures such as biopsies, injections and therapeutic implants. To aid the physician in image-guided freehand needle placement procedures, an array of assistive techniques, such as medical robots, handheld needle guides and mechanical arms have been tried out. Several research groups have investigated true three-dimensional (3D) augmented reality systems. Generally, it has been felt that robotics and 3D augmented reality devices brought significant cost and engineering complexity and thus have not been widely accepted in routine clinical practice.

Freehand needle placement, even when used with assistive techniques such as image overlay was found to be limited by requiring hand stability [Fichtinger-2005, Fischer-2007]. A need is seen to develop a training system on which physicians can hone their skills in variety of clinical situations. Contrary to casual observation, needle-based surgery can be a complex intervention. Translational and rotational motions, as well as bending and insertion forces can be combined for delicate needle control in needle-based surgery. Space and the means for desired maneuvering of the surgical device, however are extremely limited. A novice physician typically performs needle interventions under the supervision of a senior physician; a slow and inherently subjective training process that lacks objective, quantitative assessment of the surgical skill and performance.

We present a laboratory validation system, *Perk Station*, for standardized training and performance measurement under different assistance techniques for needle-based surgical guidance systems. The initial goal of the Perk Station is to assess and compare different techniques: 2D image overlay, biplane laser guide, laser protractor and conventional freehand. The Perk Station is complete with its planning and guidance software system developed on the 3D Slicer platform, a free, open source software package designed for visualization and analysis of medical image data.

2. Perk Station: System design and implementation

The overall system consists of a flat panel LCD display, a transverse plane laser, two para-sagittal plane lasers, and an electromagnetic tracker. The image overlay system shown in Figure 1 (left) comprises of a flat panel LCD display and a semi-transparent mirror aligned in such a manner that the reflection of the image appearing in the mirror coincides with the phantom behind the mirror. The transverse plane laser provides additional guidance in image overlay technique, as the image projection plane coincides with transverse laser plane. The biplane laser system shown in Figure 1 (right) consists of two para-sagittal plane lasers and a transverse plane laser. Two para-sagittal lasers are incorporated in the system to accommodate bilateral needle insertions. The transverse plane laser defines the z-plane of insertion, while the adjustable para-sagittal laser defines the in-plane needle angle. The line of intersection of the para-sagittal lasers and the transverse plane laser defines the needle trajectory. In conventional freehand technique, the transverse plane laser alone is used for guidance. The EM tracker can be used alone or in conjunction of image overlay for gestures and/or trajectory studies. By tracking the needle tip and the phantom simultaneously, we can also measure the accuracy and trajectory of needle placement in complex non-transparent phantoms. The image overlay is mounted on one side of the system and the laser guidance on the opposite side. The user can swap between the techniques simply by turning the system around. The handheld laser protractor and conventional transverse laser guidance is available on the biplane laser side, by simply turning off the para-sagittal lasers. Figure 2 shows the Perk Station in operation as the MRI image and needle trajectory is overlaid on the phantom.

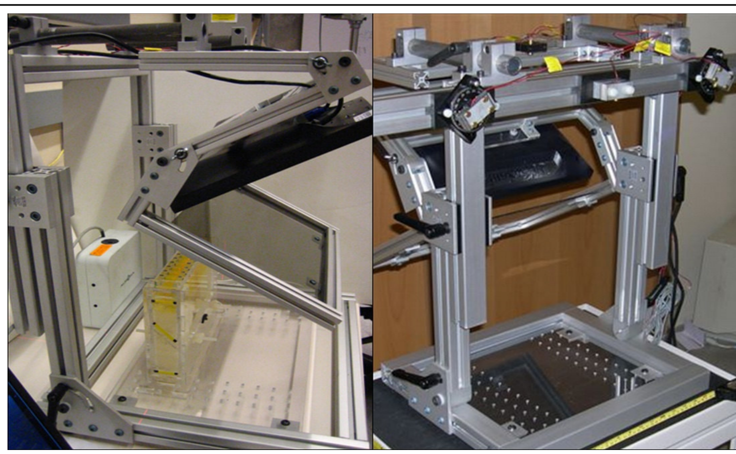


Figure 1: Image overlay system (left); Bi-plane laser guidance (right)

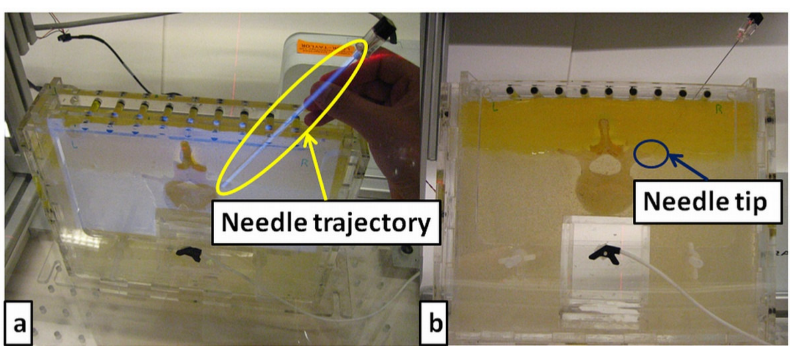


Figure 2: Needle trajectory overlaid, seen through mirror (a); Needle tip and trajectory observed in transparent phantom (b)

Author Index