Real-Time Augmented Reality for Image-Guided Interventions

Der Technischen Fakultät der
Universität Erlangen–Nürnberg

zur Erlangung des Grades

DOKTOR–INGENIEUR

vorgelegt von

Sebastian Vogt

Erlangen — 2009

Als Dissertation genehmigt von der
Technischen Fakultät der
Universität Erlangen-Nürnberg

Tag der Einreichung:	18. Februar 2009
Tag der Promotion:	05. Oktober 2009
Dekan:	Prof. Dr. R. German
Berichterstatter:	Prof. em. Dr. H. Niemann
	Prof. Dr. N. Navab

Real-Time Augmented Reality for Image-Guided Interventions

Submitted to the

Technische Fakultät der
Universität Erlangen–Nürnberg

in fulfillment of the requirements for
the degree of

DOKTOR–INGENIEUR

by

Sebastian Vogt

Erlangen — 2009

As dissertation accepted from the
Technische Fakultät der
Universität Erlangen–Nürnberg

Date of submission:	February 18, 2009
Date of graduation:	October 5, 2009
Dean:	Prof. Dr. R. German
Reviewers:	Prof. em. Dr. H. Niemann
	Prof. Dr. N. Navab

Acknowledgments

This dissertation summarizes my research activities at the Imaging and Visualization Department at Siemens Corporate Research in Princeton, New Jersey. I would like to express my deepest gratitude to Professor Dr. Heinrich Niemann and Dr. Frank Sauer for providing me the opportunity for a dissertation in the field of interventional image guidance in this corporate setting in the United States. While putting great ideas to work under Frank's direct guidance in the Interventional Imaging Group in Princeton, Professor Niemann has always given me his strong scientific support and many encouraging affirmations during my visits to Erlangen.

Taking an interventional guidance system from its first ideas all the way to pre-clinical animal trials requires a strong team and support from a variety of people. Besides Frank, I would like to thank my colleague Dr. Ali Khamene for the many fruitful discussions we had along the way. I am also very thankful for the contributions from Dr. Fred Azar and all student interns in our group who directly supported my research through user studies and software implementations. Furthermore, I would like to thank all my colleagues and friends at Siemens Corporate Research for the many valuable discussions and advise I received. Special thanks go to Professor Dr. Nassir Navab, Dr. Chanyang Xu, Dr. James Williams, and Dr. Christophe Chefd'Hotel for the discussions on augmented reality, tracking techniques, and image registration methods.

Without the strong support from my colleagues at Siemens Healthcare this research could not have progressed to the presented pre-clinical level. I am thankful for the opportunity to demonstrate the prototype system at the European Congress of Radiology, where we collected valuable feedback. A major part of the study results were collected together with our collaborators at clinical sites, such as UCLA, Brigham&Women's Hospital, Case Western Reserve University Hospitals, and Hospital of the University of Pennsylvania. Therefore, I am especially thankful for the contributions of Dr. Frank Wacker, Dr. U. Joseph Schoepf, and Dr. Gregory Rubino, and everybody involved in those pre-clinical trials. Their clinical expertise has helped tremendously to transform an engineering idea into a potential clinical device.

I would like to thank Prof. Dr. Heinrich Niemann and Prof. Dr. Nassir Navab for reviewing my manuscript and the valuable feedback I have received. Furthermore, I would like to express my gratitude to the examiners Prof. Dr. J. Hornegger, Prof. Dr. H. Niemann, Prof. Dr. N. Navab, and PD Dr. M. Lell for providing me the opportunity to defend my dissertation.

Last, but not least, the support of my family and friends was critical in my being able to finish this dissertation. I am indebted to their continued interest in the progress of my research.

Sebastian Vogt

Reality is merely an illusion, albeit a very persistent one.

—Albert Einstein

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In the realm of scientific research, chance is rarely acknowledged as a contributing factor in fundamental findings—with a few exceptions. Owing to the accidental discovery of x-rays and their properties in 1895, Röntgen's research in physics rapidly advanced the field of medicine into the completely new era of radiology, while providing the basis for some of the most important diagnostic tools and treatment procedures developed throughout the 20th century. After all, accidental discoveries have always been made by bright scientists, or as Louis Pasteur[1] coined, "In the field of observation, chance favors only the prepared mind." Prof. Röntgen was awarded a honorary Doctor of Medicine degree from the University of Würzburg and the first Nobel Prize in Physics in 1901.

Today's medical imaging technologies, such as radiography and fluoroscopy, CT and MRI scanning, ultrasonography, nuclear medicine imaging, and endoscope-type devices, provide a facet of anatomical and functional patient information to perform a reliable diagnosis and to facilitate proper treatment options. As such, medical images are not only utilized for diagnostic purposes but also for interventional guidance. The structural and functional information gained from the images and their precise mapping into the space of the patient during an intervention provides effective guidance and critical support to the performing physician during the procedure. For instance, during a brain tumor resection the neurologist's ability to completely extract all tumor tissue, while avoiding critical vessels and harming as little healthy tissue as possible, depends on proper registration between the patient's MRI data and his head as well as on an effective way to guide the neurologist toward the target structure.

[1]Louis Pasteur (1822-1895), French microbiologist, chemist, pioneer of the "Germ theory of disease", and inventor of the process of Pasteurization. *http://en.wikiquote.org/wiki/Louis_Pasteur*

	Tools	*Imaging*	*Navigation*
interventional radiologist	needle, catheter	online: x-ray fluoroscopy, ultrasound, (rarely MRI & CT)	through real-time imaging (low quality)
surgeon	scalpel	offline: MRI, CT, (ultrasound)	systems for mapping medical images to patient (neurosurgeon)

Table 1.1: A generalized view at the two main groups performing image guided interventions.

1.1 Interventional Image Guidance

For a wide variety of interventions and diseases the surgical approach has been fundamentally altered owing to the availability of patient specific imaging information in the operating room. An intervention can be categorized with respect to its invasiveness:

- *invasive*: such as open surgery or chemotherapy,
- *minimally invasive*: such as a needle biopsy or laparoscopy,
- *non-invasive*: such as virtual endoscopy or radiotherapy.

The integration of medical images into the workflow of an intervention provides the physician with information on anatomical structures and pathology of the patient that he can expect to encounter, enabling him to perform an efficient treatment and to possibly follow a minimally invasive approach. Image-guided procedures are mainly performed by surgeons and interventional radiologists. Due to the different applications their methodologies vary fundamentally and can be roughly classified as shown in Table 1.1. On the one hand, medical imaging actually permits a *radiologist* to perform minimally invasive interventions, such as needle biopsies and catheter-based interventions, guided by real-time imaging modalities, such as x-ray fluoroscopy or ultrasound imaging. On the other hand, medical imaging permits a *surgeon* to perform a surgery more targeted and effectively and therefore also enabling highly complicated interventions, possibly minimally invasive. Nevertheless, interventional radiologists are limited to navigational guidance by real-time imaging, which provides limited image quality compared to offline imaging devices, such as CT and MRI. Contrary, surgeons are limited to (high-quality) pre-operative images, mostly CT or MRI, without much navigational guidance and no imaging updates during the surgery.

In the 1980s imaging modalities were beginning to be combined with stereotactic neurosurgery techniques for the resection of brain tumors [Kel82]. *Stereotactic surgery* or *stereotaxy* utilizes a 3D coordinate system to locate small targets inside the patient and to guide a minimally invasive intervention for lesion biopsies and tumor ablations (i.e., removal), and other tasks, such

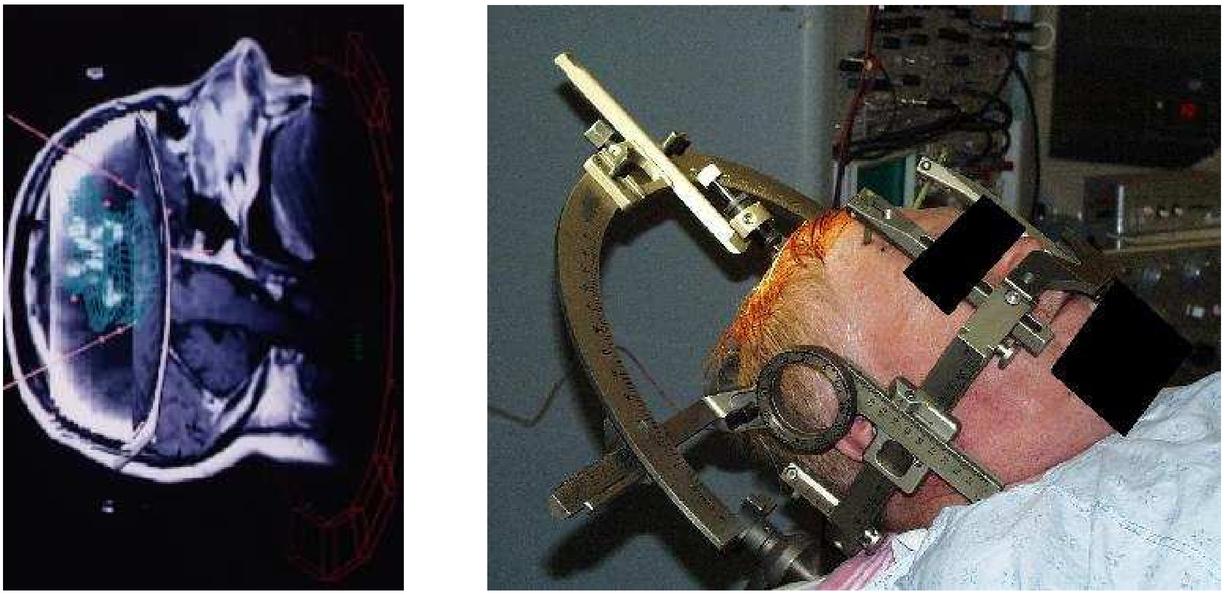


Figure 1.1: Frame-based stereotactic neurosurgery. *Left:* The medical images reveal the target structure and the stereotactic frame in the same coordinate system. *Right:* The stereotactic frame is attached to the patient's head and the surgical apparatus, which is fixed to the frame and can be moved along the circular frame, is adjusted to approach the target coordinates within the brain.

as injection, stimulation, implantation, and radiosurgery of a specific location within the patient's body. However, its applications have been mostly limited to brain surgery due to the necessity of a reliable frame of reference, such as bones with a fixed spatial relation to the soft tissue. (Nevertheless, applications for stereotactic biopsy of the female breast to sample and remove malignant tissue can be found as well.) The concept of stereotactic brain surgery is based on medical images of the patient's brain which guide the surgeon. To navigate towards the target location one of two fundamentally different techniques is utilized to create a reliable coordinate system throughout the procedure:

- *frame-based:* An external frame is attached to the patient's head that provides mechanical guidance and alignments for minimally invasive instruments.
- *frameless:* Imaging markers are attached to the patient's head and a real-time tracking device is utilized to relate instruments and pointers to this coordinate system.

In frame-based neuro-navigation a light-weight frame is attached to the patient's head under local anesthesia. Figure 1.1 shows this stereotactic frame attached to the patient's skull. Subsequently the head is imaged by CT, MRI, or x-ray angiography methods. Target locations within the images can now be extracted in spatial relationship to reference points on the frame, which

Figure 1.2: Frameless navigation system for neurosurgery. *Left:* The patient's head is fixed to the head clamp of the neurosurgical table and the pre-operatively acquired MRI images are registered to a stationary coordinate system utilizing markers that were attached to the patient's skull prior to imaging. The two-camera tracking system to localize the handheld instruments is standing behind the doorway. *Right:* The neurosurgeon can observe tracked instruments on a screen in relationship to the MRI images.

also appears in the medical images. The surgical apparatus, which is attached to the head frame, can be adjusted to the target coordinates (while being restricted to the degrees of freedom that the frame provides). Utilizing this surgical apparatus, the surgeon can now accurately approach the target within the brain. A typical example is the minimally invasive stereotactic biopsy of deep lying brain tumors. An open surgery of deep tumors within the brain poses a big challenge and would be very risky. The surgeon utilizes a stereotactic biopsy apparatus which is fixed to the head frame and adjusted along this circular head frame to target the coordinates of the tumor. Then the biopsy needle is passed from the frame through a small hole in the skull to sample tissue for diagnosis. The same technique is used to place electrodes into the brain for the treatment of movement disorders, such as Parkinson's disease.

Contrary to frame-based navigation, frameless stereotactic surgery utilizes some form of fiducial markers, which can for instance be taped to the patient's scalp prior to the CT or MRI scan. These fiducials are used later on in the operating room to register the brain images (which also contain these fiducial markers) to a stationary coordinate system on the operating table. A separate tracking system is needed to track the position and orientation of medical instruments or pointers during the surgery relative to this stationary frame. Typically a monitor screen then continuously shows a graphical representation of the tracked instruments in relation to the medical images. Figure 1.2 provides an example of a frameless navigation system. The patient's skull is rigidly attached to the head clamp of the neurosurgical table which provides a base for the coordinate system of the tracked instruments. Tracking of instruments and pointers typically works optically by means of several infrared LEDs that are attached to the instrument and which are localized in 3D space through triangulation by two or three cameras. Other tracking methods are based on passive retroreflective markers that are illuminated by a stationary infrared light source in combination with a setup of two or three cameras. Frameless neuro-navigation is an important tool for the precise approach and removal of large brain tumors due to the access and freedom it provides compared to a frame-based approach.

Commonly an *image-guided intervention* is defined as a patient encounter where medical images are obtained (during or before the procedure) and used for guidance, navigation, and orientation in a minimally invasive procedure to reach a specific target under operator control [Hal02]. The common requirements for image-guided interventions are a source of images and a real-time interactive display linked to the intervention, which also provides a means of defining the target in the context of the real 3D space (as distinguished from the abstract image space). Advantages of image-guided interventions are manifold: They are less invasive and more efficient (in time and cost), they result in fewer complications and reduce the need for rework, and additionally they tend to be the preferred approach by the public, resulting in a rapid clinical acceptance. Disadvantages are that complexity and costs may be increased for procedures that can be accomplished without medical images or images might not be updated during the procedure, opening the possibility to miss-guidance.

Image-guided interventions are often the only alternative for patients with no obvious signs of disease but positive screening results for cancer and other diseases. A reliable diagnosis for cancer, for example, can only be made and treatment started when histopathological results are available from tissue sampling. A dedicated image-guided biopsy procedure may also be combined with an administration of therapy. In the past these diagnoses had to be (and for many applications currently still have to be) performed during open surgery. Image-guidance can provide new treatment options that may not exist today. Based on the research and the proven benefits of image-guidance for breast biopsies and prostate radiotherapy many patients would prefer such an approach if it were available. Therefore there is a strong motivation to develop, refine, adapt, and extent image-guidance systems to new applications, especially to common interventions where image-guidance for a minimally invasive approach is currently not an option. Potential applications include patients with positive cancer screening tests, neurosurgery, orthopedics, vascular surgery, and general surgery.

1.2 Contribution of this Thesis

The aim of this thesis is to design, develop, and study a new approach to interventional image guidance based on augmented reality visualization—the merging of real and virtual images. Even though frameless stereotactic navigation systems, which are commercially available, currently support image-guided interventions, surgeons and interventional radiologists are restricted to looking at a monitor screen away from the actual patient, considerably complicating hand–eye coordination. A fundamentally novel augmented reality approach requires a unique combination of dedicated contributions to an array of engineering disciplines, namely systems engineering, computer vision, data visualization, and human-computer interface design. Furthermore, significant contributions to the medical sciences community have to be made based on pre-clinical studies and evaluations of the new approach. Important results here feed directly back to further advancements in the engineering of the augmented reality guidance system.

Contributions to the field of systems engineering. The complexity of an augmented reality system cannot be equated with the sum of its parts. This work, therefore, broadly reviews and categorizes all preliminary augmented reality prototype systems for interventional image guidance and serves as a first comprehensive reference in this field. A novel augmented reality setup is introduced and engineered for interventional guidance tasks, integrating pre-operative imaging, namely CT and MRI, as well as real-time ultrasound imaging. Built on a single PC platform with four independent frame-grabber boards and genlocked cameras, this system uniquely combines head-mounted stereoscopic video capture and visualization with optical head and instrument tracking in a very efficient, but yet modular, closed-loop approach. Because of its unique architecture it provides unmatched real-time performance and accuracy of high resolution augmented reality, enabling most complicated guidance tasks.

Contributions to the field of computer vision. This work advances methods for multi-camera calibration and real-time tracking. A novel head-mounted camera-triplet is described, where two cameras capture the stereoscopic view of the scene while a third, near-infrared camera with a wide-angle lens and a synchronously pulsing infrared flash is used for real-time pose estimation tasks, tracking retroreflective markers in the scene. A specialized calibration technique for the calibration of all intrinsic and relative extrinsic parameters of the camera-triplet with a dedicated calibration object is developed, taking into account the very different optical properties of the three cameras. This technique builds on established photogrammetric methods for the calibration of a single camera by means of 2D–3D point correspondences. In respect to real-time optical tracking, a unified approach to pose estimation is developed that includes an optimized cluster design of retroreflective markers, a strategy for marker identification in the tracker camera images, and a method for extrinsic camera parameter estimation based on an initial 3-Point algorithm and non-linear parameter optimizations utilizing all available 2D–3D point correspondences. This unified approach provides precise pose estimates for the camera-triplet in respect to stationary marker frames and marker clusters that are attached to hand-held tools in less than 10ms. The robustness of the tracking results is evaluated in Monte-Carlo simulations and in actual experiments.

Contributions to the field of data visualization and human-computer interfaces. The introduced augmented reality system provides a novel way to access three-dimensional datasets intuitively in direct relation to real objects and interactions. The unique combination of high-speed stereoscopic rendering of volumetric datasets (such as CT scans or ultrasound scans), head-mounted visualization, and head tracking enable the user to move around and study the seemingly stationary dataset from a variety of viewpoints. In comparison to typical screen-based visualization methods, the stereoscopic and kinetic depth cues of the proposed head-mounted approach provide a much higher degree of immersion and, thus, understanding of the data. Furthermore, the system provides the stereoscopic view of the real scene (such as the patient's body) in direct relation to the virtual data, effectively rendering real objects transparent if the virtual dataset represents otherwise invisible internal details. Interactions with the datasets, which in a screen-based solution need to be resolved through mouse movements and clicks, is given in a way that resembles interactions with real objects by means of tracked instruments. For instance, a tracked hand-held pointer provides all the means to generate arbitrary cross-sections through the 3D dataset—as simple as moving a transparency sheet through the dataset.

Contributions to applications of augmented reality. This work makes strong contributions to potential applications of augmented reality in the interventional arena. Studies are performed that show that three-dimensional graphics presented within the actual patient with stereoscopic and kinetic depth cues provide a very compelling and direct access to anatomical information for the planning and guidance of an efficient intervention. Surgeons benefit from this direct overlay of medical images onto the actual patient especially during minimally invasive surgeries. Additional navigational support is given by means of tracked instruments, in particular biopsy and ablation needles, which are visualized as virtual objects within the patient's actual body. Results of pre-clinical phantom and animal studies are presented, utilizing the introduced augmented reality system for needle placement tasks, that evaluate the benefits of the availability of pre-operatively acquired imaging information (e.g., CT or MRI) to interventional radiologists during an intervention. Increased biopsy precision, operating time, and the absence of continuous x-ray exposure through fluoroscopy are major contributions here.

1.3 Outline

This thesis is structure in the following way. Chapter 2 defines the field of augmented reality and its specific challenges when put into the realm of interventional image guidance. Its enabling technologies—medical imaging, visualization, registration, and tracking—are described in detail thereafter followed by a survey of current research on augmented reality prototype systems in the medical arena.

A novel augmented reality system, engineered for interventional image guidance, is detailed in Chapter 3. Emphasis is put on the novelty of its design and its unconventional tracking approach. High-quality stereoscopic head-mounted video see-through augmented reality puts particularly high demands on hardware and software architectures of a single PC based system. Its specific application-driven engineering and implementation is explained before the tracking methodology with a single head-mounted camera and retroreflective markers is described.

A technical evaluation of the introduced system is given in Chapter 4. The performance of the head-based tracking system as well as an optimization of marker set parameters is studied followed by an evaluation of the navigational aspects of the system by means of user studies with a tracked needle setup, also comparing the system performance with traditional screen based navigation systems.

Chapter 5 puts the AR system engineered in Chapter 3 and optimized in Chapter 4 into the context of image-guided applications, targeting surgery planning, MRI-guided neurosurgery, CT- and MRI-guided interventional radiology, as well as real-time ultrasound procedures. Each application is described based upon an adaptation of the AR system to the particular clinical setup. Extensive pre-clinical phantom and animal studies are evaluated to study the potential of the given guidance approach for future clinical use.

After providing directions for future research with the proposed augmented reality approach in Chapter 6 the thesis concludes with a summary in Chapter 7.

Chapter 2

Augmented Reality for Image Guidance

Image guidance systems help the physician to establish a mapping between a patient's medical images and the physical body. In conventional systems, a pointer or an instrument is tracked and the location visualized in the medical images on a screen separate from the patient. A new approach for image guidance is based on a visualization method that combines medical imaging information with the real view of the patient; it visually augments reality to guide the physician. The physician can see beyond the surface, the patient's body becomes transparent for him. This method has the potential to be the most direct and intuitive way of presenting the medical imaging information.

After defining the field of augmented reality, this chapter details the challenges that this technology faces especially in the arena of medical image guidance. An explanation of the different technologies to generate medical images is followed by the theoretical fundamentals of visualizing those in an augmented reality framework. Besides visualization, this framework builds upon two more enabling technologies, registration and tracking, which are described in detail thereafter. This chapter concludes with a comprehensive investigation and classification of existing approaches for augmented reality image guidance.

2.1 Definition

Although the term *Augmented Reality (AR)* has been widely used in publications throughout the last twenty years, a unique and unconditionally accepted textbook definition has not been settled yet. This is mainly due to the different technical approaches that have emerged during this time to realize an augmentation of the real world in manifold types of potential applications. There are no indications that the term Augmented Reality has been discussed before the advent or

Figure 2.1: Example for combination of virtual and real objects in a real environment as generated by an AR system. The industrial environment is augmented by a planned addition of pipes and a technical drawing from beneath the floor. *(Picture courtesy of Nassir Navab)*

outside the realm of these technical approaches. As such, Augmented Reality, in a broad sense, is understood as an alteration of the perception of the real environment by technical means. More specifically, it is a new form of interaction between humans and technology in which the user is given supplementary information about his environment, for example, computer-generated visual information via head-mounted displays. This supplementary information is context dependent—drawn from and fitted to the real object being viewed.

Rather than defining generally what Augmented Reality is, many research groups have accepted a terminology, such as in [Azu01], to define an *Augmented Reality system* as a system with the following properties. It

- combines real and virtual objects in a real environment,
- runs interactively, and in real-time, and
- registers real and virtual objects with each other.

The definition makes use of the attributes *real* and *virtual* to specify the two types of objects that are being combined. There, "real objects" denotes objects that physically exist in the environment of the user's attention. Whereas, "virtual objects" describes objects that physically do not exist but are generated in some form by the Augmented Reality system to be perceived as objects by at least one of the user's senses. Specific for an Augmented Reality system is that the combination of real and virtual objects happens in a real existing environment that the user could interact with. As such, it is intrinsic that an Augmented Reality system has to preserve this

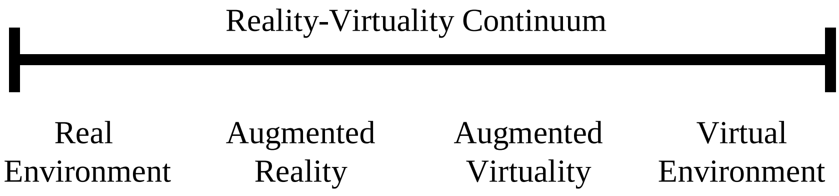


Figure 2.2: MILGRAM'S definition of Mixed Reality as a continuum between real and purely virtual environments. (Adapted from [Mil99].)

interactivity between user and real environment. To achieve this, it has to run in real-time, i.e., without a noticeable delay or perturbance of the user's perception of the environment comprising real and virtual objects. Additionally, the physical relation between the real and virtual objects has to be known to the system at any time. Therefore, it has to be able to register the virtual and the real objects with each other and keep track of the user's interactions.

Noteworthy is that this definition does not restrict itself to the sense of sight only. As such, a technologically advanced Augmented Reality system could potentially apply to all senses, such as hearing, sight, touch, and smell. Studies for multi-sensory AR and interactions, such as auditory or tangible augmentations of the real environment, can be found for instance in [Loo99, Pai05]. However, mostly all research for AR technologies has been done in the area of visualization by means of computer graphics, where computer generated objects are superimposed onto the view of the real environment.

Figure 2.1 shows an example for a seamless visual combination of real and virtual objects in a real environment as described in [Nav99b]. The view of an industrial environment is augmented with a planned addition of pipes. The user of the AR system is able to investigate if the 3D computer model of pipes fits into the existing construction and avoids possible interferences with existing parts by looking at the fused scene from different viewpoints.

In [Mil94, Mil99] a taxonomy has been introduced to describe the merging between real and virtual environments within a global framework. Computer-generated purely virtual environments, also referred to as *Virtual Reality*, on the one hand, and real environments, on the other hand, can be understood as poles at the opposite ends of a *Reality-Virtuality Continuum* as symbolized in Figure 2.2. Any world that comprises real and virtual elements can be located along this continuum. *Mixed Reality* is defined as the term to address the whole spectrum in between both poles. As such, Augmented Reality covers one part of the spectrum whereas *Augmented Virtuality* covers another. The difference between both is that the surrounding environment is real in the first and virtual in the second case.

Creating a so-called semi-immersive environment by means of an Augmented Reality system is such a fundamentally new way of presenting information to the user, that a multitude of applications for this new technology is currently being explored and investigated. Research in the field of AR tries to overcome its general as well as application specific challenges and pushes the boundaries of its enabling technologies.

2.2 Challenges

Various technologies from a facet of engineering disciplines are fundamental to tackling the challenges of an augmentation of the real world by technical means. Although an augmentation of the real world could refer to any of the user's senses, all applications so far evolve around a visual augmentation with computer graphics. Current research and prototype development of augmented reality systems draws from advances in at least three enabling technologies—visualization, registration, and tracking.

- *Visualization:* The visualization component of an AR system is responsible for the rendering of the virtual objects and the seamlessly combined presentation of virtual and real world to the user.

- *Registration:* The registration component establishes a spatial relationship between objects in the real scene, virtual objects, and the user, usually by means of parameterized models, where a subset of those parameters can change over time.

- *Tracking:* Tracking denotes the process of keeping track of the parameters over time that relate real and virtual world to each other from the user's point of view while using the Augmented Reality system.

Different applications require different approaches for these three building blocks of an AR system. Current research on AR systems concentrates on medical, manufacturing, infotainment, and military applications. Nevertheless, the fundamental challenges for building an AR system for any application are mostly the same and evolve around the three afore mentioned building blocks. The characteristics by which an AR system is evaluated are:

- *Real-Time:* An AR system needs to operate in real-time. This means that the view of the augmented world should be updated several times per second. In the movie industry it is commonly accepted that an update rate of 24-30 Hz is the minimum rate to give the viewer a feeling of continuity rather than a feeling of viewing a sequence of single images. The same applies to an AR system. Achieving a high update rate is one of the major challenges when building an AR system. Furthermore, besides the update rate for the visual cue, the real-time aspect of an AR system also comprises its system responses to user interactions. The user should have the feeling that action and reaction are in direct relation to each other.

- *Display:* The display is the central part of current AR system prototypes since it is the primary interface between the user and the AR system. The challenge is to design a display that can visualize an augmented real world to the user in a way that—while adding virtual objects to the user's view—does not deteriorate the perception of the real world. In practice a compromise has to be found that weighs the amount of alteration of the real world perception against the importance of other aspects of the AR system, such as the precision of the overlaid virtual objects and the impression of a fused augmented world. Technical means have to be found to combine the real world view with the rendered virtual world. A wide variety of display technologies are being currently evaluated for this purpose, which include head-mounted and stationary displays, stereoscopic and monoscopic devices, and optical and video techniques to blend graphics into the real environment.

- *Accuracy:* To achieve a precise visual merging of real and virtual objects in 3D space many aspects of the AR system have to be described by accurate mathematical and geometrical models. As such, the method of the display system for passing through a view of the real world to the user and its approach of overlaying virtual objects in form of 3D computer graphics have to be modeled. Registration methods have to be developed that can correlate the model of the AR display with the real world and the model of the virtual world, as well as with a tracking device which is responsible for reporting changes in the registration parameters over time. The accuracy of an AR system depends on precise models and good estimates of its model and registration parameters. Errors can be grouped into static and dynamic errors. Static errors are caused by mechanical misalignments, incorrect models and registration parameters, and errors in the tracking system, which results in ill-fitting graphics overlays even when everything, including the user, remains completely still. Dynamic errors are caused by instabilities of the tracking system or noisy estimates of its time-changing parameters, which causes jitter or visible jumps in the graphics overlay. Another form of dynamic error is introduced by latency in the AR system.

- *Latency:* Since an Augmented Reality system needs time to estimate current tracking parameters, render the view of the virtual objects and embed them into the real world view, an AR system always introduces some latency between movements in the real scene (including user actions) and its augmented visualization. This system latency translates to a perception of *lagging*, which is independent of the afore-mentioned update rate of the real-time aspect of an AR system. Currently, well designed AR systems might be able to cut system latency down to 60ms–180ms. The main difference between AR systems is the approach of how they deal with this system lag. If it is technically not possible to delay the visualization of the real part of the scene, then the augmenting computer graphics will lag behind during scene or user motions. This can be perceived as “swimming” of the graphics on the background of the real scene, which introduces, sometimes strong, visible dynamic errors to the system. A system which can delay the view of the real world, for instance by using video cameras, might be able to synchronize real and virtual world in a way that both match in space and time. In this case the lag is felt by the user for the completely fused scene, as if the whole (augmented) world were lagging behind by a split of a second. The challenge is to design a system that minimizes system latency, without compromising tracking accuracy and quality of the augmented images, and which handles the introduced lag in a way that is acceptable for the chosen application.

- *Range:* Every AR system is limited by a working range. The boundaries are defined by its construction and its tracking technology. On one side of the spectrum, mobile systems might be able to be used in many places, but a sophisticated tracking technology has to be provided to achieve satisfying accuracy. For instance a GPS (Global Positioning System) device can provide position data in the outdoors around the globe, but only with limited accuracy. This might be enough for virtual annotation tasks of the real environment in the area of infotainment, but will lead to strong misregistration errors when combining complex virtual objects with the real world view. On the other hand, stationary systems with a confined workspace can be designed around a specific application and adapted with a dedicated tracking device to provide optimal tracking results for a predefined set of tasks.

- *Acceptance:* As with every new technology, AR technology has to prove its reason for existence—which usually is very different to iterative improvements of already existing and accepted technologies. The development of a prototype, which can be studied in the planned environment of the system, especially by the users who will ultimately benefit from this new technology, is an important aspect during the research of a proper system design. Since an AR system does not automate existing tasks, nor does it control anything, but it is meant to augment a user’s vision for an application dependent purpose, the human-factor is a central part during the evaluation of such a new technology. Do the benefits of the AR system outweigh its drawbacks? Is the perception of the augmented scene strenuous to the eyes and the user? Does it help to ease or speed up tasks? Does the AR system enable handling of tasks that were unapproachable before? Are the costs appropriate? Studies that answer these and other user-related questions ultimately decide if the system is a success and research and development can continue or if the system and with it its approach fails.

Various AR projects are being pursued as the interest in AR technology is growing. Faster computers and better displays are making AR more practical and affordable. In the medical arena, under the paradigm of computer aided surgery, augmented reality systems are being researched and developed to support image guided interventions by mapping medical data directly onto the patient's body. The term *in-situ* visualization describes this scenario well: anatomical structures are being displayed at the location where they actually are. The physician can see beyond the surface, the patient's body becomes transparent for him. This is the most direct and intuitive way of presenting the medical image information. One of the most important requirements for an AR guidance system in the medical field is a very high precision for the augmenting graphics at all times. But then, compared to other AR fields, the environment is controllable, local, and confined. Besides technical challenges, acceptance factors include the possibility for sterilization of the device, the ease of use for the surgeon or interventionalist under time pressure, patient safety, and the reliability of the system.

Besides the three fundamental AR technologies of visualization, registration, and tracking, *medical imaging* is the forth enabling technology for an Augmented Reality guidance system in the medical field. To build a system that can guide a physician during an interventional procedure, the exact internal location of the patient's anatomy has to be known to the system. Medical images, such as from computed tomography, diagnostic ultrasound, and other modalities are the source for the virtual graphics that a medical AR system will use to augment the real patient and to guide the physician. Figure 2.3 illustrates those technology modules. It is the powerful combination of these four enabling technologies that will make AR a valuable tool in the future of interventional image guidance.

2.3 Medical Imaging

The twentieth century brought a multitude of inventions and discoveries that revolutionized clinical medicine. Until then the only investigative tools that were available were the microscope,

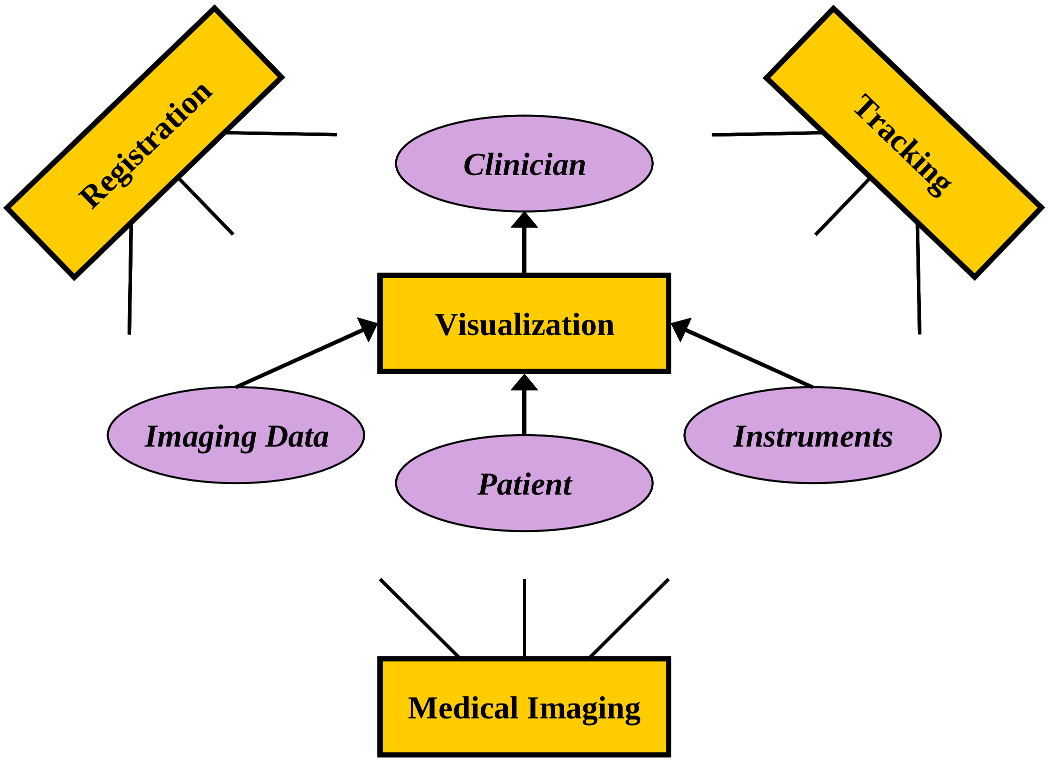


Figure 2.3: Schematic illustration of the four enabling technologies for the development of an interventional augmented reality guidance system.

the thermometer, the knife, and the stethoscope. Basically no means existed to examine and study the hidden interna of the living human body. All knowledge had been gained from centuries of careful anatomical studies of the dead and functional deficits of the injured and diseased. It was on November 8, in 1895, when Prof. Wilhelm Conrad Röntgen's discovery of the x-ray at his laboratory in Würzburg, Germany, started a revolution in medical imaging and with that a process that reunified medical science with the sciences of physics, chemistry, and engineering [Guy00]. The development of medical imaging technologies is exemplified in Figure 2.4, which shows Röntgen's first radiograph next to a volume-rendered three-dimensional view of a modern CT scan of a hand. Some basic knowledge of the technology and application of existing medical imaging devices is necessary to understand their capabilities as well as limitations and shortcomings if utilized for interventional image guidance on their own, but also to grasp their potential as an enabling technology for AR research. The following sections will give background information on state-of-the-art medical imaging technologies, which are an essential asset in developing an AR interventional guidance device.

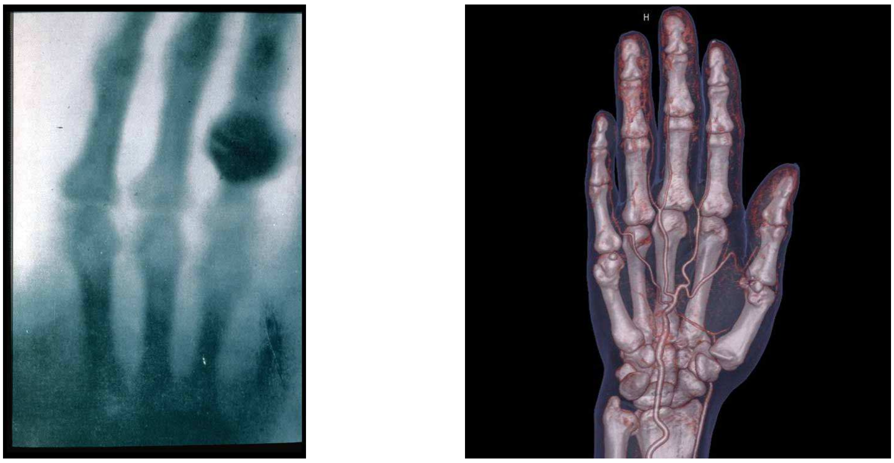


Figure 2.4: Development of medical imaging technologies. *Left:* First radiograph taken by Röntgen of his wife's hand in 1895. *Right:* Modern CT scan of a hand taken with a Siemens SOMATOM Sensation 64 scanner.

2.3.1 X-Ray Imaging and Fluoroscopy

X-ray imaging is based on the physical property of biological tissue of being partially translucent with respect to x-ray photons. The notion of the existence of photons goes back to work by Planck (1889) and then Einstein (1905), whose ideas built the fundamentals for a theoretical framework presented by Heisenberg and Schrödinger in 1926, which came to know as *Quantum Mechanics*. This name stems from the fact that the exchange of energy between atoms and light can only take place in discrete lumps. It is well known now that with the *photoelectric effect*, when monochromatic light is used to eject internal electrons from isolated simple atoms, the kinetic energy of the ejected electrons is constant, independent of the intensity of the light source. Those fundamental complex interactions could be explained based on individual encounters between particles and their exchange of energy.

The relative number of photons, emitted from an x-ray tube, that traverse a patient's body undeviated by interactions with atoms of the body is the measure that is used for medical imaging with x-rays. A well-defined beam of photons is attenuated by an amount characterized by two numbers, the Mass Attenuation coefficient for the tissue concerned and the path length of the beam within the patient. Different types of tissues, such as bone, muscle, and fat, differ in their Mass Attenuation coefficient, which eventually leads to a very specific overall attenuation of the x-ray beam when it exits the body. The x-ray photon absorption within the body is dominated by the photoelectric effect. The tissue atoms absorb energy from the x-ray photon, which leads to a disruption of the atomic structure, and charged electrons get ejected into the surrounding

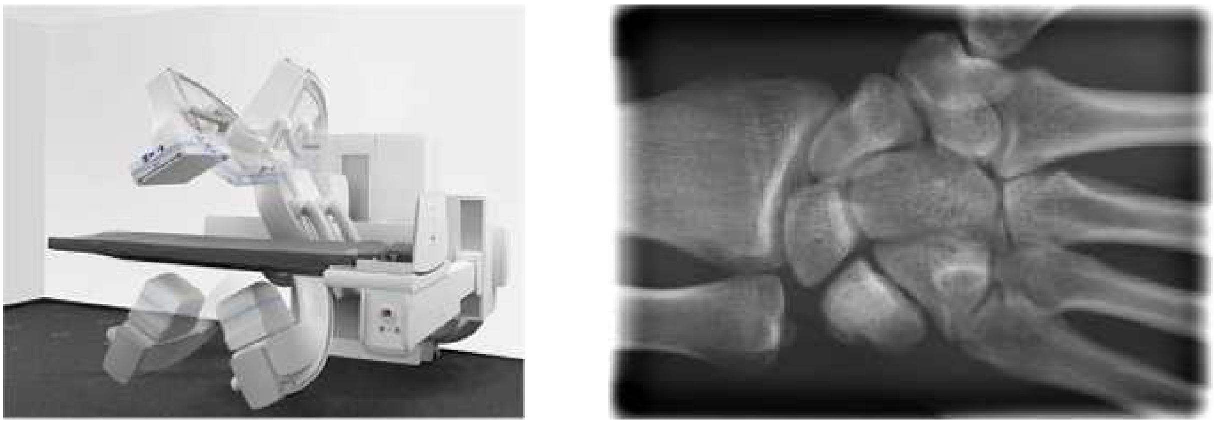


Figure 2.5: *Left:* A state-of-the-art x-ray imaging device of the C-arm type with flat-panel detector (Siemens AXIOM Artis dMP)—used for fluoroscopy, angiography, and interventional procedures. *Right:* Two-dimensional projection radiograph of the wrist.

space. Those lose their energy to the nearby tissue by creating more ions and free radicals in a cascade. It is obvious that this ionizing radiation itself poses a health risk in the long-run if applied in uncontrolled doses. But it is known that the advantages of well defined and protocolled diagnostic x-ray imaging to assess and analyze a patient's conditions outweigh those risks significantly.

A modern x-ray tube of a medical imaging device uses the same basic principle that Röntgen employed in 1895: Electrons that are accelerated in a cathode ray tube hit the target anode with a kinetic energy of 30–150keV, which is converted into heat and a broad spectrum of electromagnetic radiation that includes x-rays. In *projection radiography*, or simply *x-ray imaging*, the x-rays, which traverse the body and get attenuated depending on the different types of tissue and bone, are captured on photographic film or some kind of receptor on the opposite side of the patient. This two-dimensional projection image has a big diagnostic value. For the diagnosis of hard tissue, such as bones, a high energy photon source is used. Soft tissue organs, such as the lungs, are captured with a less-penetrating x-ray beam.

Real-time x-ray images of a patient's anatomy are obtained by *fluoroscopy*. The name stems from the first devices for that purpose, where the patient was placed in between an x-ray tube and a fluorescent screen. The room had to be darkened since the quality of those images was much inferior to that of x-ray film, especially also since the overall radiation dose to the patient has to be kept low. The development of x-ray image intensifiers in the 1950s made the use of fluoroscopy feasible for use under normal lighting conditions and to record the fluoroscopy images on camera. Modern devices for fluoroscopy employ x-ray image intensifiers and CCD cameras

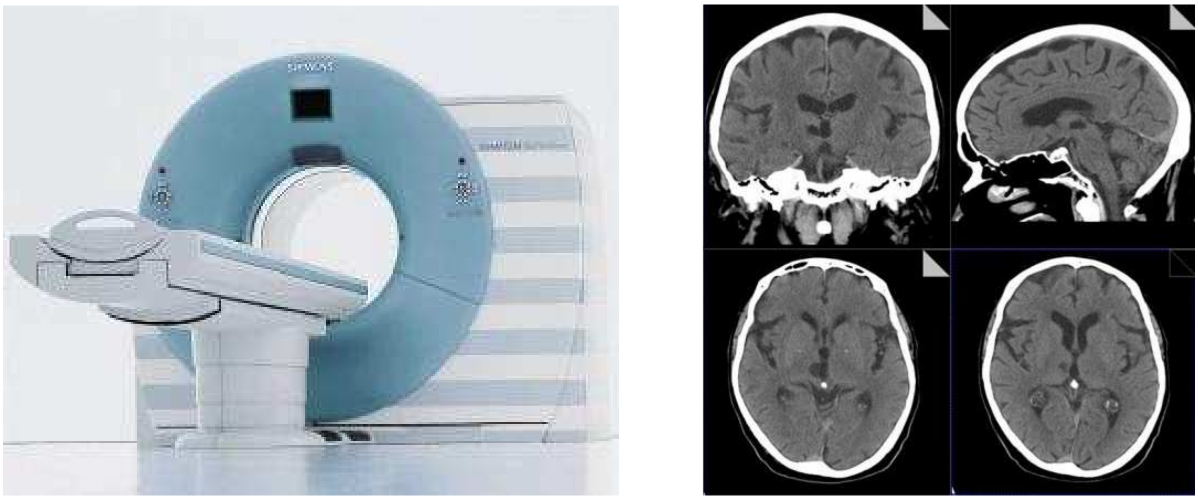


Figure 2.6: *Left:* Computed Tomography (CT) scanner (Siemens SOMATOM Definition). *Right:* Reconstructed 2D cross-sections of the head: coronal, sagittal, and two axial MPRs.

to display the real-time images on a separate monitor screen. Latest developments include the replacement of the intensifier-camera combination with flat panel detectors that are sensitive to x-rays. Figure 2.5 shows a modern x-ray imaging device, which is mounted on a C-arm and can easily be placed around the operating table during interventional and fluoroscopic procedures. Devices of this kind are especially employed for angiographic interventional procedures to visualize veins, arteries or heart chambers with x-ray imaging, where a radiocontrast agent (which absorbs x-rays) is injected into the patient's vessels. A fluoroscope is a very valuable tool during interventional procedures and guides the interventionalist with real-time updates of the patient's internal anatomy. The drawback is the exposure to radiation to the patient and the physician throughout the procedure.

2.3.2 Computed Tomography

In 1972, Godfrey N. Hounsfield invented and engineered a new medical device, a *Computed Tomography* (CT) system, which drew from the known properties of x-rays and the availability of fast computers (1979 Nobel Prize in medicine). Since then the development of CT systems has continuously progressed and computed tomography has become one of the most important tools in medical imaging. Today it is the gold standard in the diagnosis of a multitude of diseases.

The main principle behind current CT systems is an x-ray tube that rotates around the patient with a detector array on the opposite side on the circle. Projections from around the circle are recorded and used to calculate cross-sectional images through the patient's body by a method

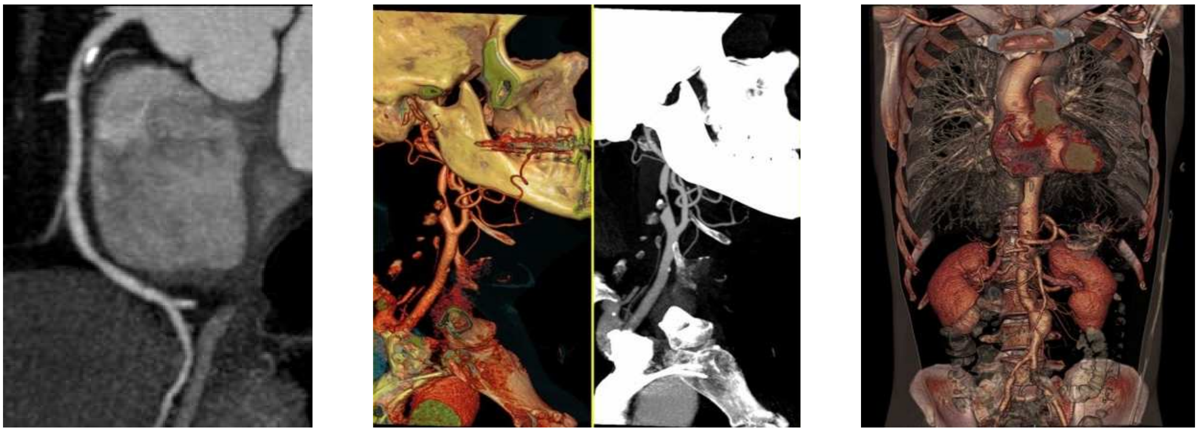


Figure 2.7: Different types of visualization of CT data. *Left:* Curved multi-planar reformatting of a coronary artery with life-threating calcification. *Center:* 3D visualization and maximum-intensity projection image of an injured carotid artery. *Right:* 3D volume rendering of abdomen.

called *tomographic reconstruction*. Whole body CT scanners (such as the one in Figure 2.6) employ a scanner table on which the patient is gradually moved through the scanner gantry to generate cross-sectional image stacks from the part of the body that needs to be diagnosed. Unlike conventional or fluoroscopic x-ray devices, CT scanners don't record 2D projection images. In its simplest form a CT scanner just contains a one-dimensional array of sensor elements along the periphery of the circle, capturing the x-rays that fan out from the tube and "slice" through the patient's body. CT cross-sectional images can only be reconstructed from multiple x-ray projections along the circle. Several cross-sectional images from along the body can be stacked together to generate a 3D volume dataset of the patient's imaged body part. Once the volume dataset is built, arbitrary two-dimensional slices through this 3D dataset in any angle can be generated by a technique called *multi-planar reformatting* (MPR). Figure 2.6 shows the most common cross-section types, here of a patient's head: *coronal*, *sagittal*, and *axial* cuts. Compared to Fluoroscopy, CT does not provide real-time images since the CT images have to be reconstructed from the multiple x-ray projections first.

CT also has many advantages over conventional x-ray imaging. CT images are not projection images and as such eliminate the superimposition of structures outside the area of interest. Furthermore, the contrast resolution of CT images is much higher, which allows to distinguish tissue types that differ by less than 1% in physical density. The use of multi-planar reformatting and 3D rendering techniques allows viewing 2D images and 3D volumes from any arbitrary direction, once the complete body part has been scanned. Figure 2.7 shows examples of different types of visualization and rendering techniques of volume data, which was acquired by CT.

CT research and development has mainly focused on scan speed, image resolution, and multi-slice scanning, where multiple slices are acquired in a single rotation by means of a detector array with multiple parallel detector rows. Combined with a spiral scanning technique (with a continuously moving scanner table), which mostly replaced the former sequential acquisition method, CT scanners have become very fast, with a total scan time of a few seconds, and deliver high-resolution volumetric datasets with isotropic voxels in the sub-millimeter range. Latest developments include the introduction of a dual-source CT scanner (example shown in Figure 2.6), which comprises two x-ray tubes and two detector arrays on the same circle, with a 90 degree angle in between.

Due to the great contrast and detail of CT images, CT scanners are used for a multitude of diagnostic assessments, involving all body parts—cranial, chest, cardiac, abdominal, pelvic, and extremities—and body scans in emergency situations. To highlight body structures, such as blood vessels or tumor tissue, which would otherwise blend in with neighboring tissue, contrast material, e.g., intravenously injected iodinated contrast, is used for certain CT scans.

Although a CT scanner is mostly used for diagnostic purposes, it still serves as an interventional guidance tool for a few procedures, either inside the gantry (*CT Fluoroscopy*) or where a *stop-and-go* approach can be utilized. An interventional radiologist would, for instance, place a biopsy needle onto the patient's body and—in an iterative way—progress the needle forward, move the patient into the scanner gantry, perform a CT scan, move the patient out of the gantry to further advance the needle in the proper direction and so forth.

2.3.3 Magnetic Resonance Imaging

The knowledge gained from the study of nuclear magnetic resonance has led to the development of a different type of scanner that works without the use of ionizing x-ray radiation. The relaxation properties of excited hydrogen nuclei in fat and water are the key elements for *Magnetic Resonance Imaging* (MRI). The first whole body scanners were build in the late 1970s. Following a sequence of developments and speed-ups, MRI found its way into clinical practice in the 1980s.

MRI is based on the interaction of static and gradient magnetic fields. The most visible part of an MRI scanner is usually its long bore hole which contains powerful magnets to create a strong uniform magnetic field in its center (see Figure 2.8). Sometimes other constructions can be found, usually with weaker magnetic fields, such as open magnets with easier access for interventional purposes. Common field strengths of MRI scanners lie between 0.3 and 3 Teslas. Stronger magnetic fields result in better image quality. The spins of the atomic nuclei of

Figure 2.8: *Left:* Magnetic Resonance Imaging scanner (Siemens MAGNETOM Avanto). *Right:* Sagittal cross-section of the head.

the tissue with non-integer spin numbers align either parallel or anti-parallel with the magnetic field. Due to the strong thermal energy of the tissue only one in a million nuclei follow this alignment, which is enough to detect a change in field, considering the vast amount of nuclei in a small volume. Exposing the tissue to pulses of electromagnetic energy (radio frequency pulses) in a plane perpendicular to the static magnetic field of the scanner causes some of the aligned hydrogen nuclei to temporarily move into a non-aligned high-energy state. After the RF excitation, those nuclei relax and realign while emitting energy. The emitted energy rate is recorded and provides information about the environment of the nuclei. The discrete Fourier transform is used to create images from the acquired data. Three different time constants during this relaxation process are of major interest, they are termed T1, T2, and $T2^*$ and describe the time it takes the spins to reach a certain state after an RF pulse. Weighting the signal by T1, T2 or $T2^*$, and no relaxation time (proton-density images) leads to the image contrast suitable to distinguish different tissue structures. For certain applications, a paramagnetic contrast agent is administered to the patient (orally or intravenously) to highlight certain anatomical structures.

An obvious advantage of MRI over CT or x-ray imaging is its use of non-ionizing RF signals to acquire images and thus its lack of harmful x-ray radiation. MRI and CT scanners can both reconstruct three-dimensional datasets or multiple two-dimensional cross-sections of the imaged body part. But MRI has a wide set of properties that can be utilized to generate image contrast (T1, T2, etc.), whereas CT images solely rely on a single x-ray attenuation attribute. Specific to the application, MRI scanning parameters can be adjusted to enhance contrast in certain tissue types. Another, but these days less restricting, advantage of MRI is its ability to generate cross-sectional images in any plane. CT images are acquired in an axial or near axial plane (by tilting the gantry). Nevertheless, the ability to acquire isotropic volume data with modern multi-slice CT scanners allows for reconstruction of 2D cross-sectional images in any plane with maximum image quality by means of multi-planar reformating. The advantages of CT are its fast scan speed (a few seconds compared to several minutes for MRI), its cheaper price (and thus its availability), in general its spatial resolution (down to 0.3mm isotropic voxel size versus 1mm with MRI) and its comfort (patients feel often claustrophobic when lying inside the MRI bore hole for several minutes or even up to an hour, depending on the exam).

Applications for MRI revolve around soft tissue exams. The superior contrast resolution of MRI compared to CT allows to distinguish pathologic tissue, e.g., a brain tumor, from the surrounding normal tissue. CT is the better tool for dense tissue exams (e.g., bone), contrast agent enhanced coronary artery imaging (due to its high spatial resolution and sensitivity to calcifications), and similar applications. Different MRI applications are targeted with specifically adjusted pulse sequences. They include *Diffusion MRI*, which is particularly good for the diagnosis of diseases like Multiple Sclerosis and in stroke patients, *Magnetic Resonance Angiography* for the detection of stenosis and aneurysms in arteries, *Functional MRI* to measure fast signal changes in the brain due to changing neural activity, and other techniques.

Because of the lack of x-ray radiation, MRI is well suited to guide interventional procedures. MRI images can be used to guide a minimally invasive procedure intraoperatively or interactively without harming the patient or the operator. To give the operator access to the patient, usually an *open bore* scanner must be utilized, which has a weaker magnetic field (in the 0.2–0.3 Tesla range). Due to their inferior image quality these types of MRI scanners are not wide-spread. Furthermore, it is important that all tools that will be used within the fringe field of the magnet are MRI compatible, i.e., non-magnetic.

2.3.4 Ultrasonography

Diagnostic ultrasound imaging, also called *ultrasonography* or just *sonography*, is a medical imaging technique that utilizes the properties of ultrasound waves to generate images of internal structures of a patient's body. This diagnostic use of ultrasound is different to the therapeutic use of ultrasound to treat patients. The latter involves different ultrasound frequencies and higher energy levels. The first use of ultrasound on humans has been reported at the Naval Medical Research Institute in Maryland in the late 1940s. First ultrasound measurements of heart activity were reported at Lund University in Sweden in 1953/54. Research at the Glasgow Royal Maternity Hospital in Scotland in the 1950s led to the first use of diagnostic ultrasound

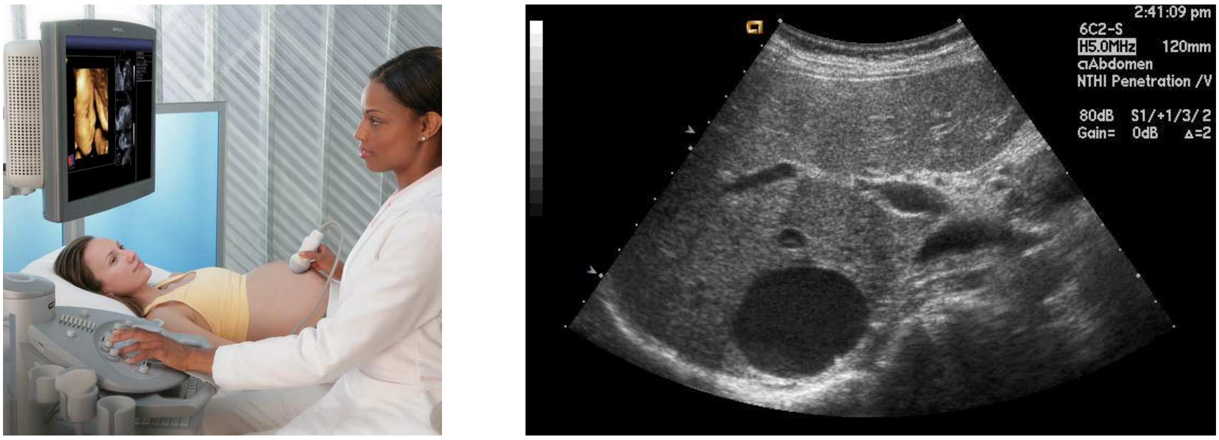


Figure 2.9: *Left:* Ultrasound machine (Siemens ACUSON Antares) in use, with the transducer probe placed on a womb of a pregnant woman. *Right:* Ultrasound image of a liver cyst.

for abdominal studies and laid the foundation for modern ultrasonography, which in the 1960s became the definitive method to study fetal growth and has since been integrated into every other area of medicine.

Ultrasonography is based on pulses of ultrasound with frequencies between 1 and 13 MHz that are generally produced by a phase array of acoustic transducers in a hand-held device, called *probe* (see left of Figure 2.9). Parts of the sound wave get reflected when it hits the interface between different tissue types (acoustic impedance). The reflection is registered by the probe as an echo. The travel time and the strength of the echo are measured to calculate the depth of the tissue interface from the probe and the difference between the two acoustic impedances. Gases or solids within the tissue reflect most of the sound pulse, which makes it difficult to see beyond those structures. Different types of transducer probes are available. A 1D phased array can be used to generate 2D images, such as in Figure 2.9 on the right. To generate 3D volume datasets, probes have been developed recently, that can either “wobble” a 1D phased array back and forth or use a more complicated 2D phased array of transducers.

Ultrasonography can effectively be used to image soft tissue and—due to its affordability—is used in many medical imaging tasks (obstetrics, cardiology, gynecology, urology, intravascular, etc.). Higher ultrasound frequencies (7–15 MHz) provide higher resolution but less penetration and are used for superficial structures (breast, tendons, muscles, ...). Internal organs (liver, kidney, ...) are imaged at lower frequencies (1–6 MHz), which provide greater penetration. A special mode is *Doppler ultrasound*, which is used to measure the speed and direction of the flow of internal fluids, especially blood, by exploiting variations of the Doppler effect.

Compared to other imaging modalities (CT, MRI, ...) ultrasonography has advantages and disadvantages. Besides being affordable, widely available, and very flexible (can be brought to the patient's bed), it has no long-term side effects and does not create discomfort. It also generates images in real-time, which allows easy adjustments of imaging parameters with direct feedback. The shortcomings of ultrasound imaging evolve around its inferior image quality and spatial resolution, as well as its difficulties in the presence of bone and gases. Ultrasound images exhibit an inherent noise pattern, known as speckle noise, which is caused by random interference between coherent backscattered waves. Obese patients are usually difficult to image due to the restricted range of penetration of the ultrasound pulses and their absorption by body fat.

Despite its cons, the real-time update of medical information and the lack of x-ray radiation make ultrasound imaging such an interesting and valuable tool for interventional guidance. The update rate of ultrasound images can be very fast; 20 images a second is not an unusual number. Even volume acquisitions with one of the latest 3D probes are fast enough to see a beating hart in three dimensions, hence referred to as 4D ultrasound imaging (3D + time). Nevertheless, the image quality of ultrasonography and the overall lack of spatial information (with a freely moving imaging probe) require a significant amount of training and expertise if used to guide interventional procedures.

2.3.5 Endoscopy

A very valuable tool for the diagnosis of certain illnesses and the guidance of minimally invasive procedures has been the *endoscope* and its variations. Unlike all afore mentioned medical imaging methods, endoscopic imaging provides direct images from within the patient's body, without introducing any type of signal or radiation into the patient. Endoscopy itself is a minimally invasive procedure, where a thin tube is inserted into the body through a natural, or sometimes artificial, body opening. Through this *scope* the interior surface of an organ can be viewed and investigated and lesions or other types of surface irregularities reported. Some endoscopes also provide the means to take biopsies. Currently the endoscope is the most frequently used tool to perform minimally invasive surgery.

The first development of an endoscope was reported at the beginning of the 19th century, but only the invention and use of electric light started to make endoscopy feasible in the early 20th century. Although selected surgeons would use and continually improve such a device throughout the first half of the 20th century, it was not until the late 1970s that a laparoscope (endoscope variant for the abdomen) would find wide-spread use for examination and surgery.

Figure 2.10 depicts a typical endoscope and images from within the small intestine. An endoscope consists of a flexible or rigid tube that contains an optical fiber systems and, depending on

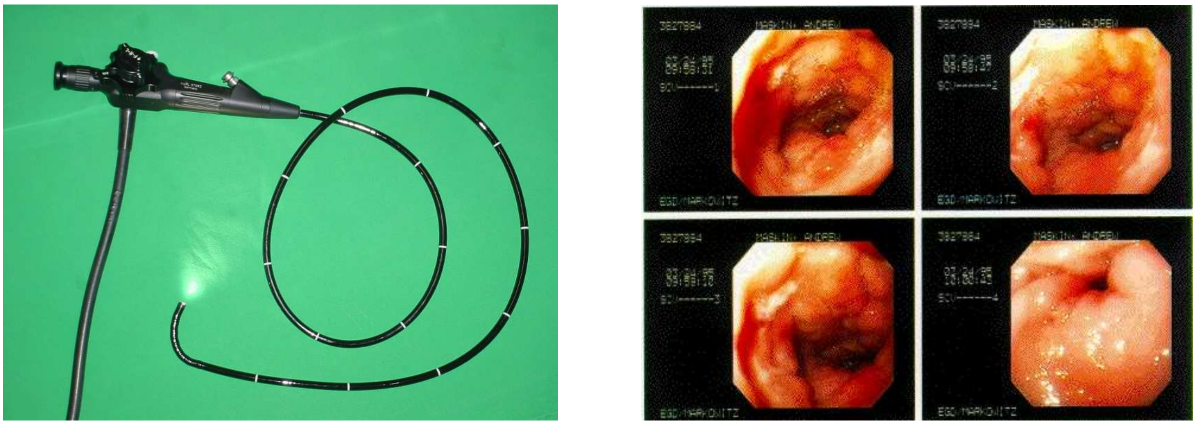


Figure 2.10: *Left:* Endoscope with flexible tube. *Right:* Endoscopic views of a duodenal ulcer (disease in the first part of the small intestine, just after the stomach).

the type, a channel for thin medical instruments. To illuminate the internal organ the endoscope also contains a method to transmit light to the tip of the tube, either by transmitting light from the outside by an optical fiber system or by small lights at the tip itself. Attached to the optical fiber system of the endoscope is a lens system that transmits the image to the viewer or to an attached camera. In the latter case (which is a typical video endoscope) the images from the camera are displayed on a separate screen.

Nowadays endoscopy is being used for a wide variety of exams and minimally invasive procedures in the gastrointestinal, respiratory, and urinary tracts, in the female reproductive system, during pregnancy, and through small incisions in the abdomen or chest, as well as for plastic surgery. Some of the more common procedures include *esophagogastroduodenoscopy* (the "typical endoscopy" of the esophagus, stomach and duodenum through the mouth), *bronchoscopy* (the lower respiratory tract through the mouth), *colonoscopy* (examination of the colon through the rectum), *laparoscopy* (abdominal or pelvic cavities through a small abdominal incision), and *thoracoscopy* (organs of the chest through a small incision). A multitude of minimally invasive procedures can be guided by endoscopic imaging. Challenges are the "keyhole" view and its spatial relationship to preoperative images, e.g., MRI or CT, and the actual patient.

2.3.6 Other Medical Imaging Techniques

Besides the listed medical imaging methods, other imaging techniques exist to extract information from within the patient's body. Nuclear medicine utilizes *gamma cameras*, *SPECT (single photon emission computed tomography) scanners*, and *PET (positron emission tomography) scanners* to measure the radiation emitted after a radioactive substance has been administered to a patient for the diagnosis of certain diseases. These types of imaging methods visualize physiological function by attaching a radionuclide to a tracer, which is the term for a chemical complex with known behavior when administered to a patient. The use of a gamma camera, SPECT, or PET scanner reveals unusual concentration of the tracer or lack of it within the body. Myocardial perfusion imaging is an example for an application of SPECT imaging. Diseased myocardium (heart muscle) receives less blood flow than normal myocardium under conditions of stress. A cardiac specific radiopharmaceutical can be used to visualize this loss of heart function with a SPECT scanner. These so called *in-vivo* tests of physiological functions also deliver spatial information and are often read alongside with CT or MRI scans, which contribute overall anatomic detail of the patient in the area of interest. The opposite to in-vivo tests are *in-vitro* tests, which are non-imaging methods, that measure samples taken from the body (blood, urine, etc.) after administration of the radioactive substance to the patient.

Another imaging method is *optical imaging*, which has been developed recently and which gives cognitive neuroscientists potentially the ability to monitor neural activity. A laser with near-infrared light is used in conjunction with detectors on the the patient's scalp to measure how the laser light gets altered. The absorption and scattering of the laser light reveals information on the concentration of chemicals and on physiological characteristics in the brain.

Methods that visualize pure physiological function are in general less applicable in interventional settings than they are for diagnostic purposes. Interventional image guidance is mostly concerned with spatial information within the patient's anatomy and fast updates during the intervention to visually guide a surgeon or interventional radiologist precisely to the target structure and at the same time to avoid other nearby critical structures within the patient's body.

2.4 Visualization

The acquisition of medical imaging information has been described in the previous section. The way of presenting this information to the clinician in a concise form that permits a precise interpretation of the patient's condition is an application dependent task. Different visualization techniques might be of interest for diagnosis, during intervention, or for medical training. Methods have been developed to visualize 2D and 3D medical imaging information on a monitor screen, which also allow interactions from the reader of the images. This section describes state-of-the-art visualization methods of medical images in form of advanced computer graphics before it details the general idea of augmenting reality visualization. The types of display technologies that are suitable for augmented reality visualization are described thereafter.

2.4.1 Medical Images

Due to the nature of each image acquisition technique, medical images are generated, handled, and viewed differently, dependent on the imaging modality.

In their most common form x-ray and fluoroscopic projection images, video-endoscopic images, and ultrasound images are two-dimensional images and are presented on a monitor screen in the form that they are acquired. Typically means are provided to change the image contrast. Endoscopy in general provides 2D color images of organ surfaces from within the body. X-ray imaging and flouroscopy provide 2D gray value projection images, where the gray-value directly decodes the attenuation of the x-ray beam by the patient's body. Ultrasound images provide slices from within the body in form of gray-value images, depicting interfaces between different types of tissue. Important for these three real-time imaging methods is that the physician typically changes directly the *acquisition parameters*, such as switching the ultrasound frequency or moving the sensor (ultrasound probe, endoscope, etc.), to adjust the visualization and to generate different views. The short feedback-loop permits this kind of workflow.

On the other hand, CT and MRI scans follow imaging protocols with a predefined set of acquisition parameters that are designed for a special task, e.g., imaging the head or the heart. Usually a stack of images is generated during such a tomographic scan and the medical technologist, radiologist, cardiologist, or physician in general is able to interact with the acquired data on the scanner console or a special post-processing machine by changing a multitude of *visualization parameters*.

Most medical imaging data is acquired in form of two-dimensional images. Although the imaging process itself is based on analog signals (x-ray attenuation, ultrasound waves, etc.), all imaging devices deliver discrete images for further processing. The extensions of these images and the range of values depend on the imaging modality and the chosen parameters during the imaging process. A two-dimensional CT image is typically defined as a function:

$$f(\boldsymbol{x}) = h \quad \text{with} \quad \boldsymbol{x} \in \{0, 1, \ldots, 511\}^2 \quad \text{and} \quad u \in \{-1024, -1023, \ldots, +3071\}\,, \qquad (2.1)$$

where $\boldsymbol{x}$ denotes the two-dimensional pixel coordinates within the image and u the CT value, which is measured in *Hounsfield units (HU)* and decoded in 12-bit integer numbers. The higher the x-ray attenuation, the higher is the CT value at a particular image pixel. Bones exhibit very high, soft tissue and water medium, and air very low CT values. Similar definitions for two-dimensional medical images can be found for all imaging modalities.

Due to the wide range of Hounsfield Units in a CT image (12 bits, i.e., 4096 levels) a special

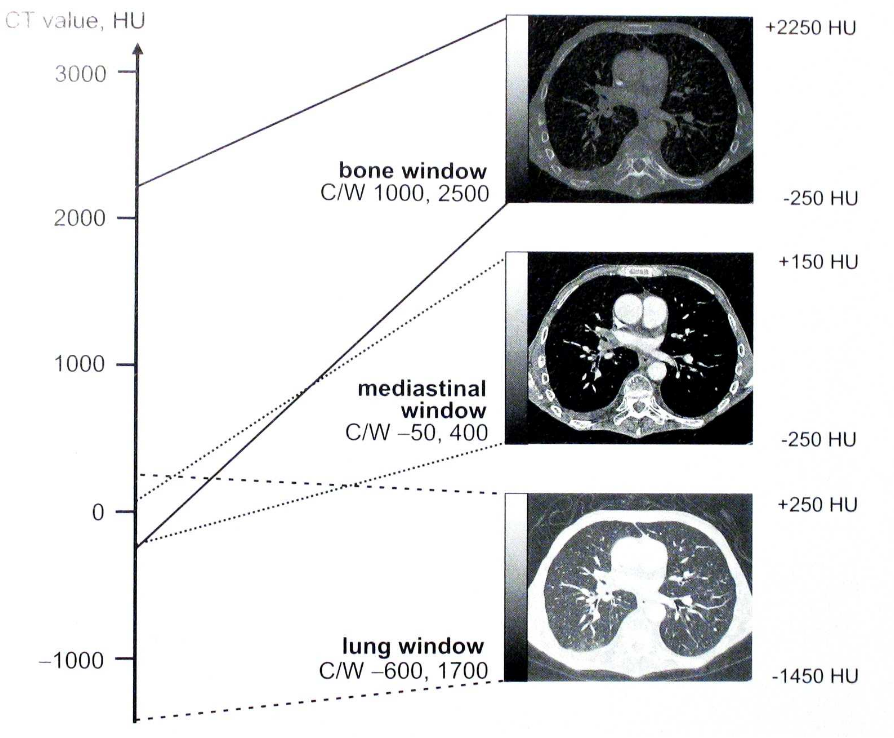


Figure 2.11: Example for windowing of CT images. The range of interest of CT values is mapped by a ramp function to a complete gray-value scale. This windowing function is defined by its center and width (C/W). Different presets are shown for a CT scan of the chest. *(Picture courtesy of Willi A. Kalender [Kal00].)*

mapping function is used that maps only an interval, which is of particular interest to the reader of the images, to a complete gray-scale. This process is called *windowing*, where the user defines the center and the width of the window. Therefore this windowing function $t(u)$ is usually defined as a ramp function

$$t(u) = v \quad \text{with} \quad u \in \{-1024, -1023, \ldots, +3071\} \quad \text{and} \quad v \in \{0, 1, \ldots, 255\}\,, \tag{2.2}$$

where v denotes a gray value of an image pixel on the screen. Values above the chosen window will be displayed as white and the values below the window as black. Since different parts of the anatomy provide very different Hounsfield Units, this windowing function provides means to maximize contrast for the tissue type of interest. Special presets are available, such as bone windows, mediastinal windows, or lung windows. Figure 2.11 exemplifies this windowing technique on a 2D cross-sectional CT image of the chest.

A windowing function can easily be implemented in form of a look-up table in any modern

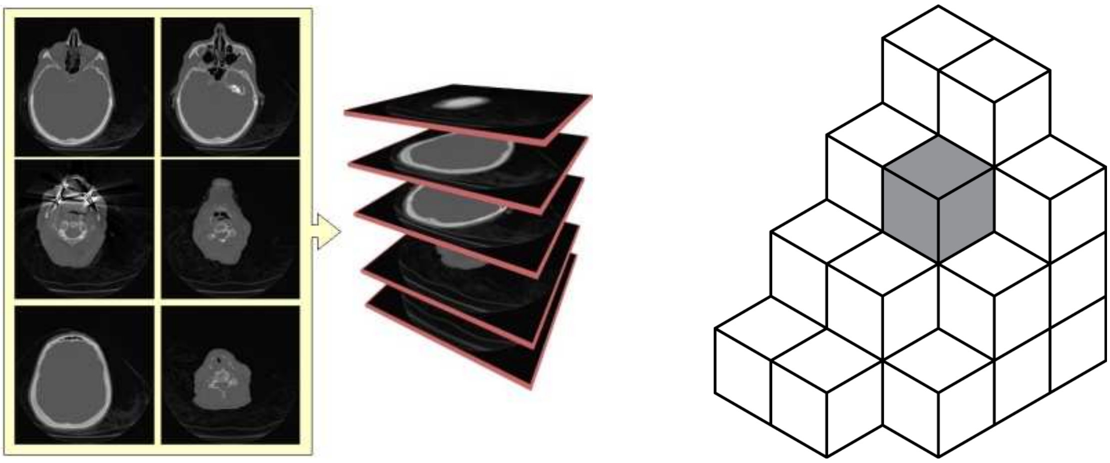


Figure 2.12: *Left:* A stack of axial CT slices of a patient's head is being arranged into a volumetric dataset. *Right:* Illustration of a set of voxels, with a single voxel in gray.

graphics hardware, without the need of converting the 12-bit dataset to an actual gray-value image. During the rendering process the graphics processor invokes the look-up table to transform the value from the dataset into a gray-value. This also permits fast and interactive changes of the windowing function by a simple update of the look-up table by the user.

2.4.2 Rendering of Volumetric Medical Images

CT and MRI scanner provide 2D cross-sectional images along the imaged body part. The clinician is able to browse through this stack of images to perform a diagnosis or to support an intervention. Although radiologists are trained in reading volumetric medical information in form of series of 2D images, for instance by browsing through a stack of axial CT images, arranging them into a three-dimensional volume dataset provides the means for better application specific visualizations and user–data interactions. Figure 2.12 illustrates the concept. A stack of adjacent (or partially overlapping) CT slices is arranged and fused into a single three-dimensional dataset. The volume element is called a *voxel* and represents the modality specific imaging values, such as Hounsfield units, on a regular grid in three-dimensional space. Contemporary CT and MRI scanners offer data acquisitions with isotropic, or near-isotropic, resolutions, i.e., the thickness of the acquired 2D slices is similar to the dimensions of a single pixel in the 2D images. This is beneficial for the generation of volume datasets with isotropic voxels.

One form of examining 3D medical datasets is by using a technique called *multi-planar reformatting (MPR)*. Arbitrarily oriented 2D slices through the volumetric dataset can be generated by interpolating their pixel values from the dataset voxels where the cut-plane intersects. This

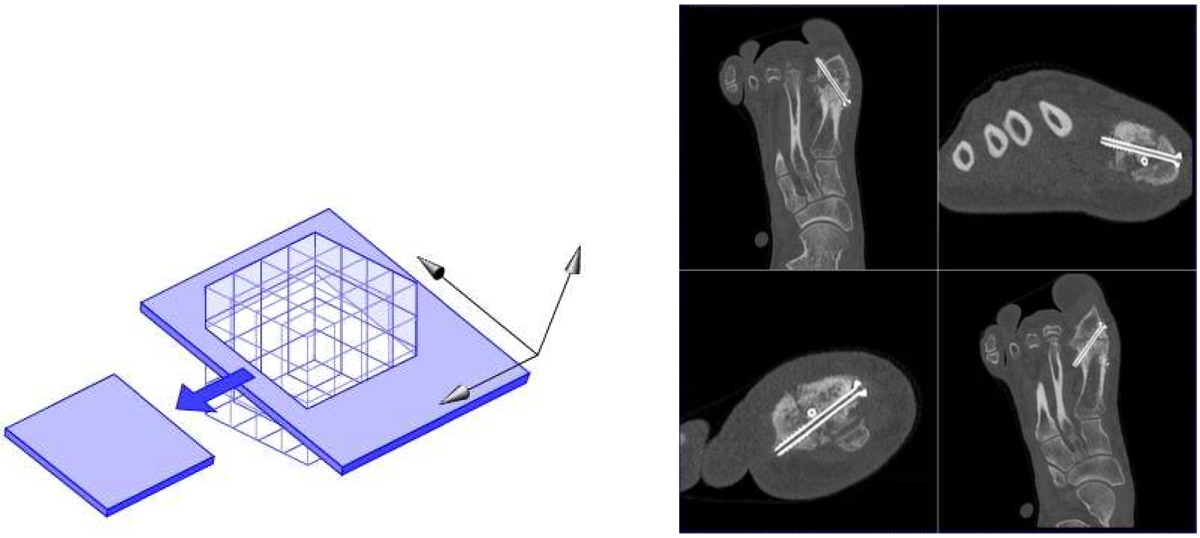


Figure 2.13: Illustration of multi-planar reformating. *Left:* An arbitrary 2D slice is being generated from a voxel-based 3D dataset. *(Picture courtesy of SGI.)* *Right:* An isotropic volumetric CT dataset of a patient's foot is being examined by generating different MPR slices through the 3D volume, which are all aligned with the patient's bone-implanted screws.

method is illustrated in Figure 2.13. The generation of MPR slices provides the means to the clinician to align the visible 2D information with anatomical landmarks, such as vessels, bones, or implants.

A possible method for implementing a fast generation of MPR slices is by using the capabilities of modern graphics hardware to store volumetric datasets in form of 3D textures directly on the graphics board. The GPU (graphics processing unit) is then able to quickly access all necessary information for further visualization tasks. To render an MPR slice, 3D texture coordinates are assigned to each corner of a planar 2D patch, which represents the MPR slice and is visualized on the screen. During the rendering process the GPU generates 3D texture coordinates for each pixel of the 2D patch by linear interpolation of the bordering texture coordinates. Eventually, the GPU determines the pixel values of the MPR slice by interpolating voxel values from the 3D dataset at the generated 3D texture coordinates for each rendered pixel of the 2D patch. The result is that the desired MPR cross-section is generated by calculating proper 3D texture coordinates from the cut of the 2D plane with the volumetric dataset. This method can be well combined with the afore mentioned look-up table implementation of a windowing function, where the interpolated voxel values are directly transformed into gray-values during the rendering process.

A visualization technique that incorporates more or less all available information from the

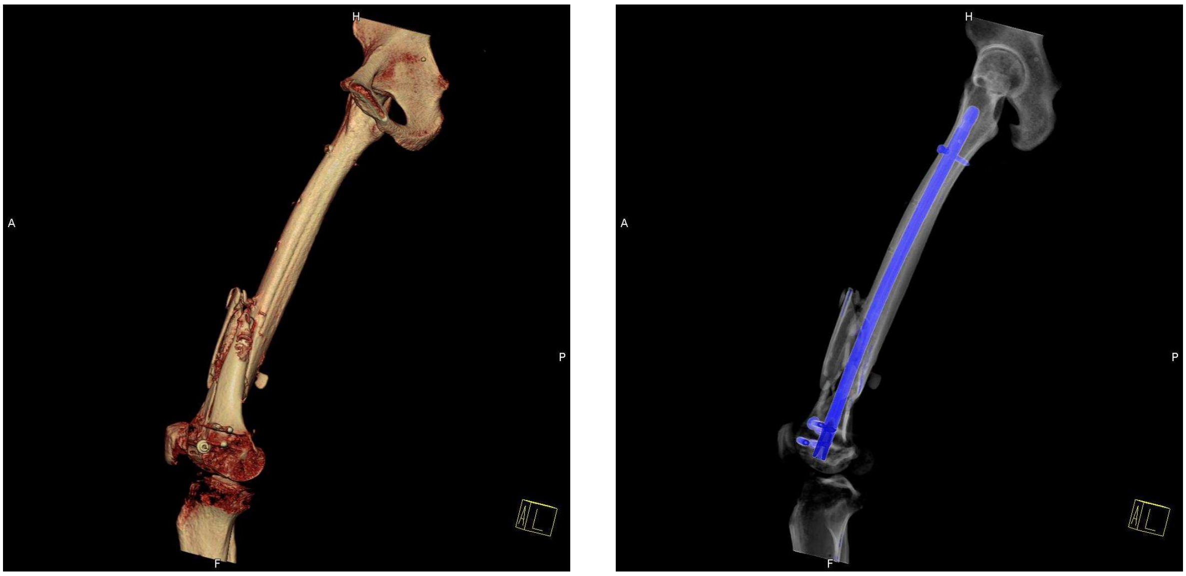


Figure 2.14: Volume rendering of a fractured thigh bone, showing the effect of different transfer functions. Both images were rendered from the same dataset and the same viewing direction. Only the transfer function has been changed to reveal different structures of the injured bone.

medical 3D dataset is called *(direct) volume rendering*. It displays a 2D projection of a 3D discretely sampled dataset. Unlike indirect volume visualization methods, such as surface extraction and its visualization, a direct volume renderer in general invokes each voxel of the dataset. Firstly, it requires a mapping of each voxel value to an opacity and color value, which is usually given as a quadruple, RGBA, where R, G, and B stand for the primary color components red, green, and blue, and A for a transparency (alpha) value. This mapping or *transfer function* $t(u)$ transforms imaging data values u, such as Hounsfield Units, into vectors $\boldsymbol{c}$, denoting the RGBA quadruple:

$$t(u) = \boldsymbol{c} \quad \text{with} \quad u \in \{-1024, -1023, \ldots, +3071\} \quad \text{and} \quad \boldsymbol{c} \in \mathbb{R}^4, \boldsymbol{c} \in [0,1]^4 \,. \tag{2.3}$$

Although t can be an arbitrary table, in most applications it is constructed out of four piecewise linear functions, such as trapezoid functions, one for each color and the alpha component. Secondly, the direct volume renderer composes all RGBA values along the viewing direction to create a single 2D projection of the volume.

Choosing different transfer functions and adjusting their parameters allows the clinician to hide and emphasize different parts of the patient's anatomy. Transfer functions have been developed that aim to imitate human flesh and bone colors, based on the Hounsfield Units of a CT scan. An example for a volume rendered fractured bone is shown in Figure 2.14. Whereas one

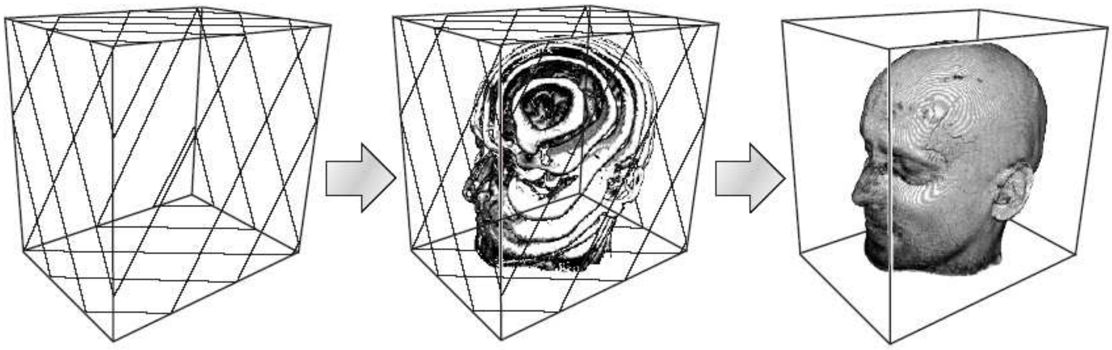


Figure 2.15: Decomposition of a volumetric dataset into viewport-aligned polygon slices for the purpose of 3D texture-based volume rendering. *(Picture courtesy of Christof Rezk-Salama [RS02].)*

transfer function might reveal important structures on the joints and the bone surface, another one renders the bone very transparent and reveals whether the titanium bone-implant has been placed correctly inside the bone. It should be noted that in this example both transfer functions render soft tissue (muscle etc.) around the bone with alpha values equaling complete transparency.

Several rendering techniques exisit that perform the composition of the RGBA values of the volume into a single 2D projection [RS02]. The most intuitive way is simply *ray casting*, where for each pixel of the 2D projection image a ray is cast into the 3D volume along the viewing direction. The RGBA values along the ray are blended with each other and a single color value is calculated for the pixel.

Methods and specialized hardware have been developed that speed up the process of volume rendering and achieve an interactive visualization for the user. One of the fastest rendering techniques is based on the 3D texture capabilities of modern graphics cards. Figure 2.15 and Figure 2.16 illustrate the concept. Similar to the generation of MPR slices, cuts through the volume are generated that are all orthogonal to the viewing direction. The GPU is very fast in rendering 2D patches on top of each other that are all textured from the same 3D dataset by specifying 3D texture coordinates for the corners of each patch. The main task is to generate a set of viewport-aligned (i.e., orthogonal to the viewing direction), slices through the 3D dataset and to render those from back to front. The renderer invokes the correct combination for each pixel of the 2D image buffer with the new patch that is rendered on top of it. The final, composed image will have the same look as a ray-casted image. As has been illustrated in Figure 2.16, the renderer can be configured to generate either a parallel or a perspective projection of the volume. This type of hardware-accelerated volume rendering can be very fast, where the user is able to

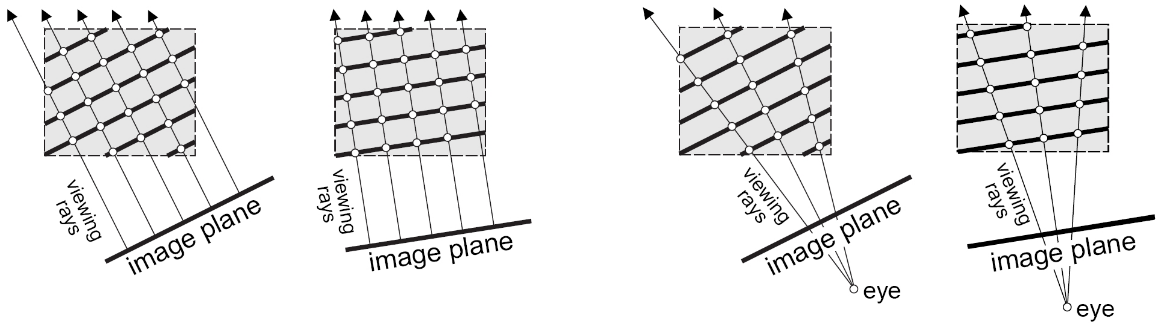


Figure 2.16: Illustration of volume rendering with viewport-aligned slices. *Left:* Parallel projection. *Right:* Perspective projection. *(Pictures courtesy of Christof Rezk-Salama [RS02].)*

change the viewing direction with interactive speeds. It is mostly limited by the size of graphics memory which must hold the volumetric dataset.

Besides the described direct volume rendering techniques, indirect techniques exist to visualize volumetric medical data. Indirect visualization methods of volumetric data rely on an extraction of boundaries of regions within the data in form of surface meshes. This type of data segmentation is performed in a preprocessing step. The actual visualization of the mesh is done by polygon rendering techniques. Modern graphics hardware provides capabilities to load triangulated surface data into the graphics memory and to render the 3D object as wireframe models or shaded surfaces, with a variety of color, transparency, texture, and shading properties.

Many automatic and semi-automatic methods exist for the segmentation of medical data. Common ones include the extraction of isosurfaces with the *Marching Cubes* algorithm or segmentation methods based on *active contour* and *level set* approaches. The extraction of isosurfaces, i.e., equal values within the volume data, is of interest if a special type of tissue, for instance tumor tissue, needs to be segmented from the dataset. It usually exhibits a certain imaging value, such as the same Hounsfield Units in a CT scan. Segmentation also permits the clinician to perform quantitative analyses, for instance of the growth of a tumor in a follow-up scan, by calculating volume or area measurements.

Although direct volume rendering is a much more common visualization method of 3D medical data due to its transfer-function induced flexibility, the segmentation of data and its representation as triangulated surface meshes is still of interest in different clinical applications. Often a combined visualization of direct volume rendering and segmented surface data is desirable. A clinician might want to manually outline a structure like a brain tumor and then visualize its surface within the volume-rendered head of the patient.

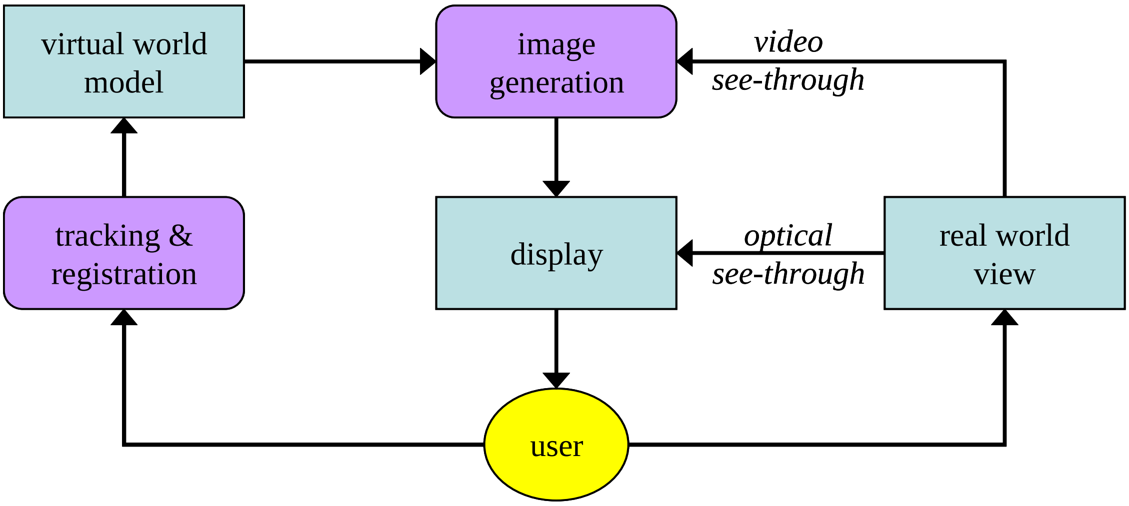


Figure 2.17: System model for an AR system. The tracking and registration information is used to render views of the virtual objects that are aligned with the view of the real world. Virtual and real worlds are blended with optical or video methods and displayed as a an augmented reality view to the user.

2.4.3 Graphics and Real Scene Fusion

Traditional visualization of medical imaging information is restricted to a monitor screen. For the purpose of augmented reality image guidance different types of displays and visualization methods need to be investigated. As has been stated earlier, besides rendering of the virtual objects, the visualization component of an AR system is responsible for the seamlessly combined presentation of virtual and real world to the user. Figure 2.17 illustrates the concept of blending virtual objects into the view of the real scene with an augmented reality system [Vog00, Vog01]. In general, the tracking and registration components (Chapter 2.5 and 2.6) of the AR system provide essential information to the visualization component to render views of the virtual objects that are spatially aligned with the view of the real world.

Two different types of fusing virtual and real images have been investigated in the realm of AR: video mixing and optical combination [Baj92, Rol00]. Those two fundamentally different approaches are also referred to as *video see-through* and *optical see-through* augmented reality. They differ in the way they present the view of the real world to the user. Video-based methods capture the view of the real environment by one or two cameras as live recorded video streams. Usually a computer then merges the virtual graphics into the video streams and the augmented result is almost instantaneously displayed to the user. For instance, a head-mounted display (HMD) can be equipped with a pair of cameras that capture the real scene and feed the video streams into a computer for augmentation and display by the HMD. Optical methods, on the other hand, provide a direct view of the real environment to the user. An optical image of the virtual graphics is generated and appears within the visual field of the user. This can be achieved, for instance, by head-mounted half-transparent displays through which the user sees the environment overlaid with the graphics on the screen.

Both augmented reality approaches have advantages and disadvantages [Azu97, Rol00]. The advantages of an optical approach over a video approach are:

- *Simplicity:* The blending of virtual images into the real scene by means of half-transparent mirrors is simpler and cheaper than the use of video cameras. Besides the video cameras, real-time processing of video streams requires more powerful computers. Also the digitization of the camera images usually adds additional delay time. Furthermore, optical approaches do not have to deal with distortion of the real scene caused by the camera optics of video see-through systems. Overall an optical see-through system can be build less complicated than a video-based system.

- *Resolution:* In a video see-through system the resolution of the real scene and the virtual objects are both limited by the resolution of the display device. An optical approach permits the user to see the real scene with the full resolving power of the retina, which currently is much higher than the resolution of available display devices. The overlaid graphics part, nevertheless, is limited by the display device.

- *No eye offset:* Since the user's view of the real world is provided by video cameras in video-based systems, the user must adjust to the viewpoint of the cameras, which in general does not coincide with the actual viewpoint of his eyes. An optical see-through system that combines real and virtual by means of half-transparent displays, provides a direct view to the real scene without any viewpoint offset.

The advantages, on the other hand, that a video see-through system offers over an optical see-through system are as follows:

- *Flexibility in composition strategies:* Fusing virtual and real worlds also requires to block light from real objects that are behind the virtual location of virtual objects. Optical see-through systems in general are not able to completely block out incoming light. Because of that the overlaid graphics appears ghost-like in optical see-through systems. Furthermore, matching the brightness of real and virtual objects is also an advantage when having the real view captured on video. Optical systems exhibit contrast problems in very bright or dark environments. Altogether, video see-through systems have the capability to achieve a more compelling augmented reality view.

- *Synchronization of real and virtual views:* Optical see-through systems offer an instantaneous view of the real world, but the virtual view is delayed due to the delays caused by tracking and rendering. This causes a temporal mismatch that becomes very apparent in dynamic scenes or with fast viewpoint changes. In the contrary, video see-through has the capability of delaying the video of the real view until tracking and rendering processes finish and to synchronize both worlds in time.

- *Additional registration strategies:* The digitized images of the real scene provide a valuable source for the registration of real and virtual worlds. An objective image based calibration routine can be employed to register the coordinate system of the virtual world with the real world by means of camera calibration techniques. Optical see-through systems need to rely on user input for such registration, i.e., in most cases the user has to specify corresponding points in virtual and real worlds to achieve a proper alignment. This makes the registration subjective and very susceptible to errors. Moreover, the registration might need to be repeated for every new session. The video see-through system can supply reliable registration parameters in an advanced laboratory setting independent of the users.

Both optical and video approaches exhibit problems in different ways with regard to focus. Video see-through delivers a combined AR image in a single focal plane defined by the monitor screen(s). However, due to the camera lens, parts of the real world might not be in focus, whereas the graphics part is sharp at all locations due to the assumption of a pinhole camera model during graphics rendering. An advanced visualization scheme might be able to tackle this issue by introducing lens blurring according to a depth-of-field model derived from the properties of the camera lens.

A typical optical see-through system displays the virtual image at a fixed location behind a half-transparent panel. This causes problems for the user when trying to focus on real objects behind the panel that are at a different distance than the virtual image. As such it may not be possible to view virtual and real objects clearly at the same time, although in 3D space they might meant to be close to each other. This can be a major perceptual problem in optical see-through displays that utilize half-transparent panels or displays.

For many tasks the correct perception of depth is fundamental when navigating in three dimensions. Augmented reality visualization has to take this into consideration. Virtual environments were studied in this respect and many of the results can be applied to the field of augmented reality. The human visual system uses physiological and psychological cues to interpret depth [Oko77]. The *physiological depth cues* are binocular parallax, motion parallax, accommodation, and convergence.

- *Binocular parallax:* Since both eyes perceive the environment from slightly different locations, the images are slightly different. This difference is called binocular parallax. For medium viewing distances this is the most important depth cue, since the human visual system is very sensitive to these differences. It is possible to achieve a sense of depth using binocular parallax even if none of the other depth cues are available.

- *Motion parallax:* Depth can be perceived by head movement. Similar to binocular parallax, motion parallax extracts depth information from images that were sensed by the visual system shortly after each other (with the same eye) from slightly different locations.

- *Accommodation:* The tension of the muscle that changes the focal length of the eye lens is called accommodation. Objects from different distances to the observer's eye are brought into focus this way. Accommodation is a weak depth cue and is only effective in very short distances (2 meters) in combination with other depth cues.

- *Convergence:* The difference in the direction of both eyes that occurs when objects in short distances are being watched while both eyes tilt slightly inwards is called convergence. It is effective in the near range of up to 10 meters.

Two of the physiological depth cues, binocular parallax and convergence, require stereo vision. In contrary, the *psychological depth cues* only require mono vision. They are: retinal image size, occlusion, texture gradient, linear perspective, aerial perspective, and shades and shadows.

- *Retinal image size:* The size of the image of a known object on the retina is utilized by the brain to estimate its distance from the observer.

- *Occlusion:* Objects that occlude each other partially reveal relative depth amongst those objects. Occluded objects are father away than the occluding ones.

- *Texture gradient:* Surface texture on objects becomes more detailed the closer the objects are. Especially objects that reach from near to far reveal a strong sense of depth this way.

- *Linear perspective:* Lines that extent into the far distance, such as parallel lines that meet on the horizon, reveal depth by providing perspective.

- *Aerial perspective:* Objects in a far distance often look hazier due to the humidity in the air. This is a depth cue used in outdoors scenarios, such as mountains in different distances.

- *Shades and shadows:* Objects that cast shadows on other objects are closer to the light source. This provides a sense for the distance between different objects to the observer.

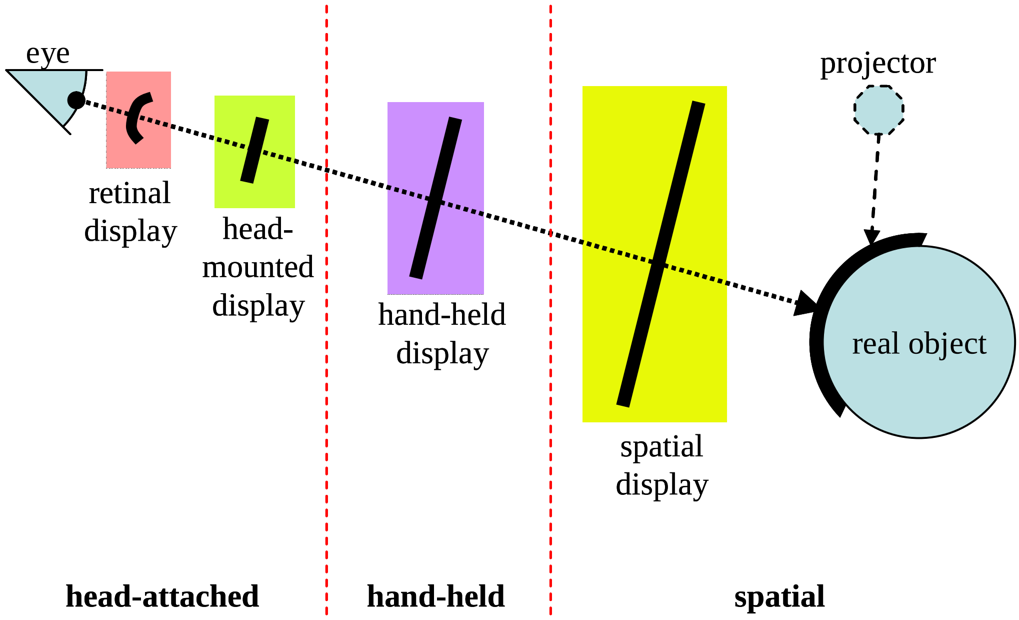


Figure 2.18: Illustration of the different types of augmented reality displays.

When merging virtual and real objects in an AR system, the visualization component has to address these physiological and psychological depth cues that influence the perception of the fused scene. The proper rendering of the virtual graphics, such as 3D medical datasets, the way of seamlessly combining real and virtual objects by means of video or optical see-through approaches, as well as the proper type of display device, e.g., support of stereo or motion parallax, are the three technological challenges of the visualization component of an AR system. The next section will detail display device options for AR systems.

2.4.4 Augmented Reality Displays

Augmented reality systems require specialized display devices [Bim05, pg. 71], [Fuc99, Azu01]. The display device is directly linked with the chosen augmented reality approach, either based on optical or video merging. As such, different AR applications demand different display technologies. The requirements for see-through properties and depth perception contribute to this decision.

Figure 2.18 illustrates the different places where an augmented reality image can be located with respect to the user and the environment. The spectrum reaches from the retina of the observer's eyes, to the surface of the real object itself. Display devices that are attached to the head are worn by the user and usually stay aligned with the user's eyes throughout the session.

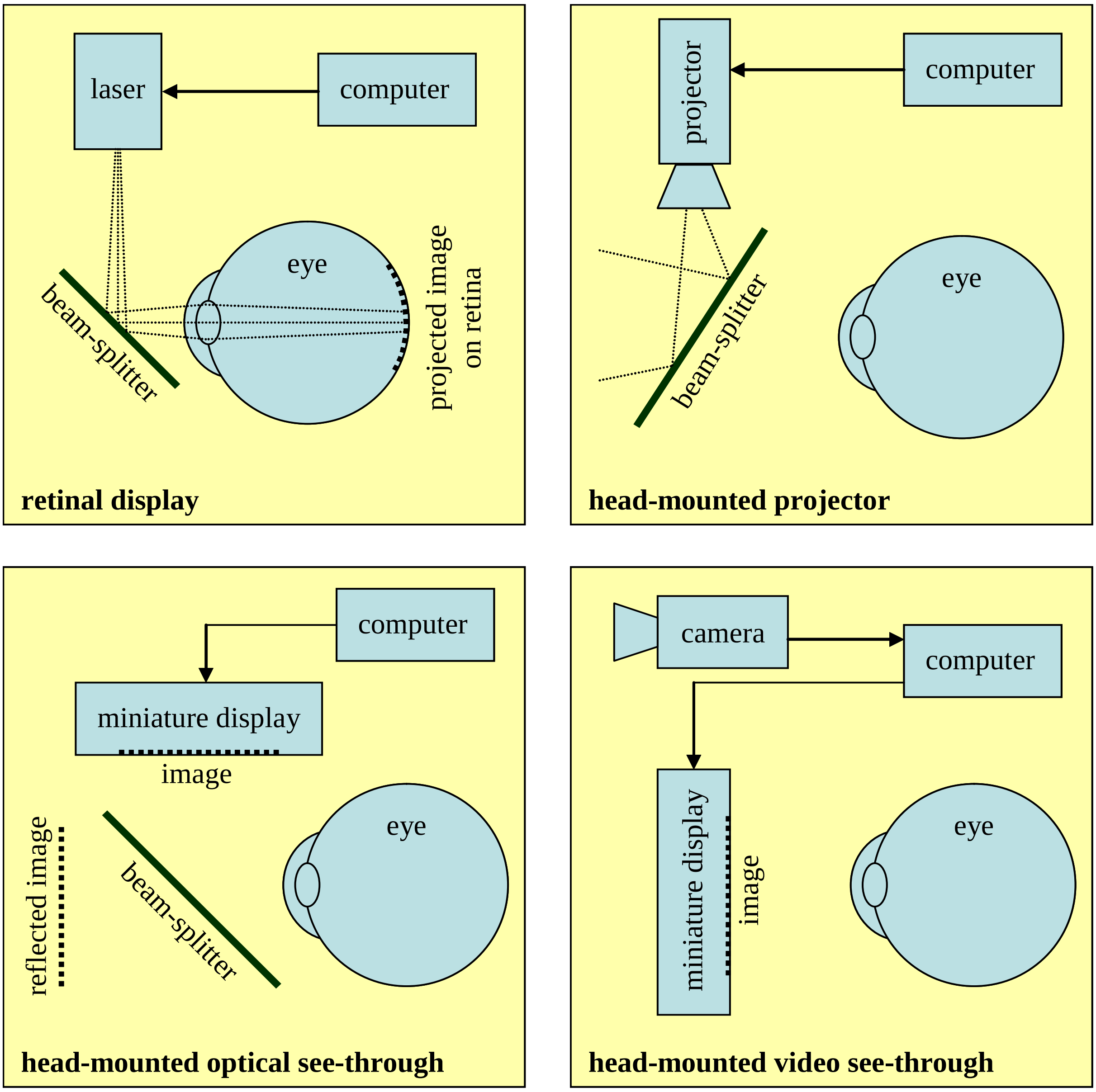


Figure 2.19: Illustration of the basic concepts of head-attached displays.

Other displays are spatially aligned with the real objects which are being augmented. Whereas the alignment of hand-held displays to the user's eyes and the real environment usually changes continuously throughout the session. Registration and tracking—two of the enabling technologies of AR—have to resolve the alignment problem of virtual and real worlds with regard to the display device and the user.

Head-attached displays can be classified into three groups: retinal displays, head-mounted displays and head-mounted projectors. Figure 2.19 illustrates their function.

A low-power laser is used for *retinal displays* to scan modulated light directly onto the retina of the human eye, without the need for any (miniature) screen or monitor. A beam-splitter in front of the eye permits the person to view directly the real environment while the laser from the top is reflected frontally into the eye. Compared to screen-based displays such a device produces brighter images with higher resolution and possibly a wider field-of-view. However, currently only monochrome versions (red laser) are available and technically feasible. Furthermore, only monoscopic versions exist, whereas stereoscopic vision is required by many AR applications to support proper depth perception.

Head-mounted displays (HMDs) are currently the most common display devices in augmented reality applications. Different versions exist that support optical see-through AR and others video see-through AR. Although the latter typically in form of prototype versions by a few research groups. Perhaps the first AR system was based on an optical see-through HMD, developed by Ivan Sutherland in the 1960s [Sut65, Sut68]. In optical see-through HMDs optical combiners are used to fuse the direct view of the real world with the reflection of a miniature display that is mounted above the half-silvered mirror. Some versions exist that make use of transparent LCD displays instead of the beam-splitter/display combination. Video see-through HMDs mount the display directly in front of the eye. One or two cameras are attached to the HMD (for instance on top) and are directed into the typical viewing direction when the HMD is placed on the head. The live video streams are transfered to the computer, augmented with graphics, and displayed on the displays in front of the eyes. Advantages of optical and video see-through HMDs is their availability as stereoscopic versions. Compared to, for instance, retinal displays the field-of-view of HMDs is more limited due to the displays and optics.

The frustum of miniature projectors is redirected with a mirror beam combiner in case of a *head-mounted projective display (HMPD)*. The redirected images are projected onto retro-reflective surfaces in front of the user of the HMPD. Retro-reflective material has the property to reflect light back along its incident direction, which results in bright images perceived by the user. To avoid the need for retroreflective material in the user's environment, *projective head-mounted displays (PHMDs)* utilize head-mounted projectors that project the images onto a regular ceiling above the user. Two half-silvered mirrors in front of the users eyes are used to reflect the stereo image on the ceiling into the user's eyes. This is similar to the optical see-through approach of HMDs, whereas the images now come from the ceiling instead of two miniature displays above the beam-splitters. HMPDs and PHMDs address problems of regular HMDs, such as to provide a wider field-of-view and a better handling of accommodation and convergence issues of optical see-through HMDs. They aim to combine advantages of projective displays with the advantages of traditional HMDs. However, the strong dependence on room lighting conditions and the need for a ceiling or retroreflective coverage of the environment make those displays more suited for virtual reality applications than AR.

Hand-held displays generate augmented reality images on devices such as Tablet PCs, PDAs, or even cell phones. The view of the real world is captured by integrated miniature cameras, augmented with graphics, and displayed on the screen of the device. Thus, most of potential AR applications with hand-held devices employ the video see-through approach. Due to the limited computing power of these devices, limited screen size, inferior camera optics, and a difficult registration and tracking situation, AR based on commercially available hand-held devices mostly targets the infotainment sector. In contrary, as emphasized in Chapter 2.2, medical applications require systems with a high precision and reliability to be able to be used successfully. Nevertheless, Chapter 2.7.4 will demonstrate a prototype of a hand-held augmented reality device for medical purposes. This system, however, utilizes optical see-through with a special mirror combination that avoids the need for registration and tracking. As such, it is possible to build very specialized hand-held AR devices to support augmented reality guidance in the medical arena.

In the contrary to head-attached and hand-held displays, spatial displays are detached from the user's body and are integrated into his environment. The existing approaches for augmented reality visualization differ in the form they augment the environment by means of optical see-through, video see-through, or projector-based direct augmentation. In combination with, for instance, polarized or shutter glasses, these approaches can also provide stereoscopic viewing to support a good sense of depth perception. In this case the spatial display provides two different images for each eye, which are synchronized with the goggles that the user is wearing.

Spatial video see-through displays usually employ a regular monitor screen which displays the augmented video image. This type of AR provides a low degree of immersion, usually provides a small field-of-view (restricted by the monitor size), a reduced resolution of the real world, and gives an impression of remote viewing, rather than providing a see-through feeling. Nevertheless, it is one of the most cost efficient approaches to AR and permits good blending options and the synchronization of motion between the fused real and virtual worlds.

The images that are generated by *spatial optical see-through displays* are aligned within the user's environment. A typical approach is to mount a big transparent screen in front of the real objects that are meant to be augmented. The screen visualizes the virtual objects while the user sees the real scene behind it. Advantages of this approach are the direct view of the real scene and the more stable calibration since the screen is aligned with the environment that it augments. The manipulation of objects behind the screen can become cumbersome with this approach, though. Other shortcomings of optical see-through approaches often apply here as well, such as the focus problem, the synchronization of real world motion and virtual objects, and the contrast issues in regard to occlusion of real objects by virtual ones.

Projector-based direct augmentation of the real environment is a type of spatially aligned augmented reality that tries to overcome limitations of other approaches. One or multiple projectors that directly project graphics onto the surface of the real objects, can potentially cover a wide field-of-view and do not exhibit the focus problem of traditional optical see-through technologies. Nevertheless, new problems arise, such as the casting of shadows by physical objects within the frustum of the projector, restrictions of the graphics by the physical surfaces that are present, multi-user issues in case of stereoscopic projections, and blurring of the projector image in case the physical objects stretch over a wider range of depths. Nevertheless, due to its appealing direct augmentation, projector-based approaches that handle the mentioned limitations can lead to good AR results.

In the medical context different approaches to augmented reality have been investigated, based on optical and video see-through techniques, with head-mounted displays and spatial displays, in both monoscopic and stereoscopic implementations. Chapter 2.7 will survey the existing prototypes. As described earlier, visualization is one of the enabling technologies of AR. The augmented reality visualization of medical imaging information also requires registration and tracking techniques that support the proper alignment of virtual and real objects on the chosen AR display.

2.5 Registration

Visualization of medical images with an Augmented Reality system requires a system component that is responsible for the spatial alignment between the medical images and the real view of the patient. In general, the registration problem is one of the most fundamental problems limiting Augmented Reality applications [Azu97].

2.5.1 The Registration Problem in AR

To create an illusion with an AR system that real and virtual worlds coexist, the objects from the two worlds must be properly aligned with respect to each other. Applications, such as interventional image guidance, demand accurate registration at all times. For instance, if the visualization of a tumor within a patient is slightly off then the misaligned virtual representation will guide the surgeon during a needle biopsy to a wrong location. He will miss the tumor and most likely harm surrounding tissue. Eventually the biopsy will fail. For the acceptance of an AR system accurate registration is essential in many applications.

The spatial alignment problem has been known to the virtual reality community before. Registration problems exist in virtual environments, but they are much harder to detect than in augmented reality environments. In virtual reality applications, such as surgical simulators, the user only sees and interacts with virtual objects. Registration errors result in visual-kinesthetic and visual-proprioceptive, but never visual-visual conflicts. Since the visual system is much more sensitive than any of the other human sensory systems, visual-visual conflicts are the most noticeable. For instance, a user wearing a virtual reality HMD might hold up a real pointer in his hands, which he sees in form of a virtual pointer in his virtual environment. If this virtual pointer is displayed in a slightly wrong location, e.g., 1cm to the right, the user of a pure virtual environment might not notice that. In an augmented reality setting the real pointer and virtual pointer will be visible and the 1cm difference in location will look very disturbing to the user.

In general, an AR system needs to generate views of virtual objects with the same viewing parameters as the view of the real objects has been taken at the same point in time. Therefore, the registration task in AR can be broken down into two sub-tasks:

1. spatial registration of the virtual objects with the real objects in 3D and
2. calibration of the view of the eye/camera (optical/video see-through) to the real objects.

The first item is responsible for the generation of a coordinate mapping for each virtual object (e.g., a CT dataset of the brain or a wireframe model of a needle) to the coordinate system of its real counterpart (e.g., the patient's head or a biopsy needle). The second item is necessary to derive viewing parameters (e.g., camera parameters of a video see-through system) from the way each of the real objects is viewed that can be applied to the generation of a view of the virtual objects with exactly the same parameters. Eventually those two views will be merged by optical or video means. It should be noted that in stereoscopic AR visualization the second step has to be performed for each eye or camera.

Not all of the parameters of this registration stay constant over time. For instance, the location of a hand-held biopsy needle with respect to the operator's eye or the video camera will change with every hand motion. Keeping track of these parameter changes is the task of the tracking component of the AR system. This means, instead of re-evaluating the complete set of registration parameters, usually indirect or direct means are employed that allow a fast update of only the parameters that changed. More details about tracking will be given in Chapter 2.6.

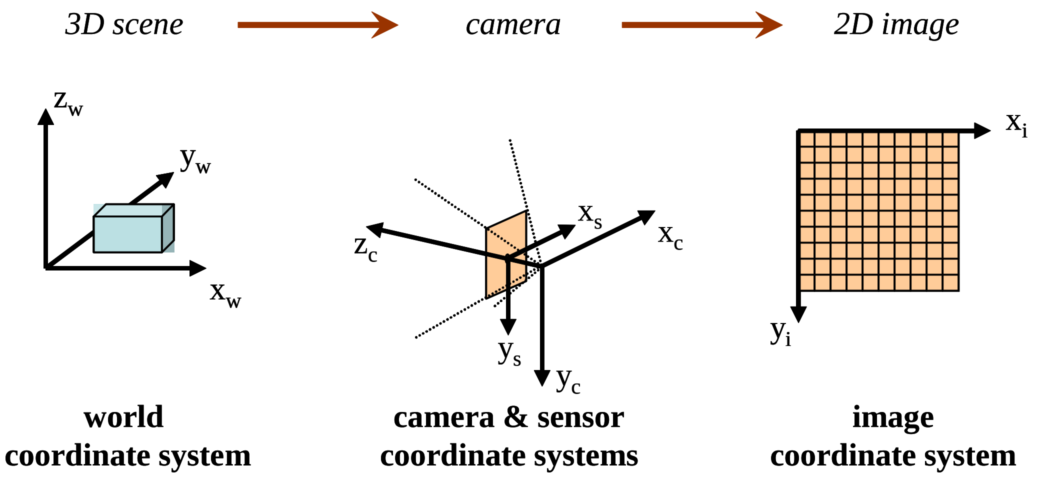


Figure 2.20: Coordinate systems that describe the common pinhole camera model and its relation to an object-centered world coordinate system. This model can be used to describe and parameterize the virtual camera of an AR system.

2.5.2 Parameters of a Virtual Camera Model

The virtual objects that augment the real scene have to be rendered by means of a virtual camera. In a video see-through AR system this virtual camera should resemble the properties of the real camera as much as possible to achieve correct overlays. In an optical see-through AR system the virtual camera should properly model the chosen eye/display combination. As mentioned in the previous section, one of the two sub-tasks of the registration component of an AR system is a system calibration to determine viewing parameters which can be utilized for a parameterized model of this virtual camera. The following will describe a common mathematical camera model with its parameters, which has been used in the fields of computer vision and computer graphics.

Figure 2.20 illustrates the coordinate systems that are relevant when modeling the transformation of 3D scene points to points in the final 2D image. The *world coordinate system* and the *camera coordinate system* are right-handed three-dimensional cartesian coordinate systems that are spatially aligned with objects in the 3D scene and the camera, respectively. The z_c-axis of the camera coordinate system describes the optical axis of the camera. The *sensor coordinate system* is a two-dimensional cartesian coordinate system and describes points on the screen or sensor of the camera, which is aligned parallel to the z_c-plane of the camera coordinate system. While all the mentioned coordinate systems usually follow metrical measurements, the *image coordinate system* is a pixel-oriented two-dimensional cartesian coordinate system. The following list describes the mathematical formulations of all addressed coordinate transformations [Vog01].

1. *Transformation of world to camera coordinates:* The coordinate transformation of an object point with given world coordinates $\boldsymbol{p}_\mathrm{w} = (p_\mathrm{wx}\ p_\mathrm{wy}\ p_\mathrm{wz})^\mathrm{T} \in \mathbb{R}^3$ into the camera coordinate system can be described by an affine transformation:

$$\boldsymbol{p}_\mathrm{c} = \boldsymbol{R}\boldsymbol{p}_\mathrm{w} + \boldsymbol{t} \qquad \text{with} \quad \boldsymbol{R} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}, \quad \boldsymbol{t} = \begin{pmatrix} t_\mathrm{x} \\ t_\mathrm{y} \\ t_\mathrm{z} \end{pmatrix} \tag{2.4}$$

   by means of an orthonormal rotation matrix $\boldsymbol{R} \in \mathbb{R}^{3\times 3}$ and a translation vector $\boldsymbol{t} \in \mathbb{R}^3$. Since the camera coordinate system is aligned with the camera, the resulting camera coordinates $\boldsymbol{p}_\mathrm{c} = (p_\mathrm{cx}\ p_\mathrm{cy}\ p_\mathrm{cz})^\mathrm{T} \in \mathbb{R}^3$ of the object point are dependent on the position and orientation (*pose*) of the camera with respect to the environment.

2. *Transformation of camera to sensor coordinates:* The usual camera model in computer graphics and computer vision is the *pinhole camera* model [Fol97, Nie90]. Due to its convenient way of modeling by means of a *perspective projection*, it is preferred over all other types of models that, for instance, involve more realistic optics-based models of thin or thick lenses. The idealized assumption is that the camera has a screen in a distance f behind a miniature hole (the optical center) where exactly only one ray from each 3D scene point passes through and generates an image point on the screen. The distance f is called the *focal length* of the camera. Furthermore, since the generated image would be inverted due to the crossing rays at the optical center, the model assumes that the screen is *in front* with the distance f of the optical center. The perspective projection of a 3D scene point with camera coordinates $\boldsymbol{p}_\mathrm{c} = (p_\mathrm{cx}\ p_\mathrm{cy}\ p_\mathrm{cz})^\mathrm{T} \in \mathbb{R}^3$ onto the 2D screen is described by the equation

$$\boldsymbol{p}_\mathrm{s} = \begin{pmatrix} p_\mathrm{sx} \\ p_\mathrm{sy} \end{pmatrix} = f \begin{pmatrix} p_\mathrm{cx}/p_\mathrm{cz} \\ p_\mathrm{cy}/p_\mathrm{cz} \end{pmatrix}, \tag{2.5}$$

   with $\boldsymbol{p}_\mathrm{s} = (p_\mathrm{sx}\ p_\mathrm{sy})^\mathrm{T} \in \mathbb{R}^2$ describing the coordinates of the projected image point on the screen or camera sensor. If f is given in millimeters then the sensor coordinates will be calculated in millimeters as well, with the center of the coordinate system at the intersection of the optical axis with the sensor/screen.

3. *Transformation of sensor to image coordinates:* The transformation of a 2D point to pixel coordinates that are given in sensor coordinates is dependent on the distance between the pixels on the sensor. Furthermore, the center of projection, where the optical axis of the camera intersects with the sensor (in pixel coordinates), influences this transformation. It

is given by:

$$\boldsymbol{p}_{\mathrm{i}} = \begin{pmatrix} s_{\mathrm{x}} & 0 \\ 0 & s_{\mathrm{y}} \end{pmatrix} \boldsymbol{p}_{\mathrm{s}} + \begin{pmatrix} c_{\mathrm{x}} \\ c_{\mathrm{y}} \end{pmatrix} , \tag{2.6}$$

where $\boldsymbol{p}_{\mathrm{i}} = (p_{\mathrm{ix}} \; p_{\mathrm{iy}})^{\mathrm{T}} \in \mathbb{R}^2$ describes the resulting image coordinates in pixels. The factors $s_{\mathrm{x}}, s_{\mathrm{y}} \in \mathbb{R}$ are scaling factors, which can be different for both dimensions. $1/s_{\mathrm{x}}$ denotes the distance between the centers of the pixels on the sensor in $\mathrm{x_s}$-direction and $1/s_{\mathrm{y}}$ in $\mathrm{y_s}$-direction. The center of projection is defined by the pixel coordinates $c_{\mathrm{x}}, c_{\mathrm{y}} \in \mathbb{R}$.

The described camera model is based on the principle of a pinhole camera. In the context of a given world coordinate system, the complete model is parameterized by two types of parameters:

- *Extrinsic camera parameters:* $\boldsymbol{R}, \boldsymbol{t}$ and
- *Intrinsic camera parameters:* $f, s_{\mathrm{x}}, s_{\mathrm{y}}, c_{\mathrm{x}}, c_{\mathrm{y}}$.

It should be noted that if no metric information about the camera sensor is given during camera calibration, the focal length f cannot be treated independently from the scaling factors $s_{\mathrm{x}}, s_{\mathrm{y}}$. This becomes clear when (2.5) gets substituted into (2.6). In this case s_{x} is simply set to 1. Then f describes the focal length in a unit that is based on the width of a pixel element on the sensor and s_{y} is the ratio between the physical height and width of a pixel.

A minimum set of parameters of the virtual camera is advantageous when those parameters have to be determined from the real view of the AR system to achieve matching viewing conditions for real and virtual worlds. Calibration techniques, which will be addressed in the next section, help in finding those parameters. It is known that an orthonormal 3×3 matrix describing rotations in 3D space only has 3 degrees of freedom. For the purpose of modeling the virtual camera with a minimum set of viewing parameters it is important to also describe those rotations with a minimum set of parameters—ideally three. The following forms of minimal parameterization of 3D rotations are common:

- *Three rotations around coordinate axes:* A common form of describing a rotation $\boldsymbol{R}$ is by a sequence of three rotations about the axes of the coordinate system by three so called *Euler angles* (e.g., in [Ant00, pg. 180], [Jäh97, pg. 192], [Fol97, pg. 215]). The first rotation is by an angle ϕ about the z-axis, the second is by an angle $\theta \in [0, \pi]$ about the x-axis, and the third is by an angle ψ about the z-axis again. This presentation can be written as $\boldsymbol{R} = \boldsymbol{R}_{\mathrm{z}}(\psi)\boldsymbol{R}_{\mathrm{x}}(\theta)\boldsymbol{R}_{\mathrm{z}}(\phi)$, where $\boldsymbol{R}_{\mathrm{z}}(\psi)$, $\boldsymbol{R}_{\mathrm{x}}(\theta)$, and $\boldsymbol{R}_{\mathrm{z}}(\phi)$ denote rotations around the z-, x-, and z-axis, respectively. It should be noted that this type of parameterization exhibits singularities, making numerical optimizations difficult or instable in certain cases [Hor99].

- *Rotation around an axis in $\mathbb{R}^3$:* Computer graphics applications often describe three-dimensional rotations by a counterclockwise rotation through an angle ω about an axis in $\mathbb{R}^3$, which is given by an arbitrary unit vector $\boldsymbol{u} = (u_\mathrm{x}\ u_\mathrm{y}\ u_\mathrm{z})^\mathrm{T} \in \mathbb{R}^3$ with $\|\boldsymbol{u}\| = 1$ (e.g., in [Fol97, pg. 227], [Ant00, pg. 181]). The transformation from $(\boldsymbol{u}, \omega)$ to $\boldsymbol{R}$ is given by the *Rodriguez* equations and the inverse transformation can be solved by singular value decomposition, since $\boldsymbol{R}\boldsymbol{u} = \boldsymbol{u}$ (all points on the rotation axis are invariant with respect to the rotation) is a typical Eigenvector problem.

- *Rotation representation by a quaternion:* Rotations can also be represented by a unit quaternion $\boldsymbol{q} = (w, \boldsymbol{v}),\ w \in \mathbb{R}, \boldsymbol{v} \in \mathbb{R}^3, \|\boldsymbol{q}\| = 1$ (e.g., in [Wei07]). With respect to the previous axis/angle representation the quaternion representation is defined as: $\boldsymbol{q} = (\cos(\omega/2),\ \boldsymbol{u}\sin(\omega/2))$. A rotation of a point $\boldsymbol{p} \in \mathbb{R}^3$ can then easily be described by quaternion multiplication and conjugation: $(0, \boldsymbol{p}_\mathbf{rot}) = \boldsymbol{q} \cdot (0, \boldsymbol{p}) \cdot \bar{\boldsymbol{q}}$, where $\boldsymbol{p}_\mathbf{rot} \in \mathbb{R}^3$ describes the new point coordinates after rotation. Furthermore, if necessary, a rotation matrix $\boldsymbol{R} \in \mathbb{R}^{3\times 3}$ can be calculated from the quaternion $\boldsymbol{q} = (w, \boldsymbol{v}) = (w, x, y, z)$ by:

$$\boldsymbol{R} = \begin{pmatrix} 1-2y^2-2z^2 & 2xy-2wz & 2zx+2wy \\ 2xy+2wz & 1-2x^2-2z^2 & 2yz-2wx \\ 2zx-2wy & 2yz+2wx & 1-2x^2-2y^2 \end{pmatrix} . \tag{2.7}$$

When modeling the virtual camera according to real optics of, for instance, a camera of a video see-through AR system, oftentimes the pinhole camera model itself does not suffice, due to optical distortions of the real camera lenses and optics. In those systems it is common to model optical distortion as a separate step in the afore-mentioned sequence of transformations from world coordinates to image coordinates. The most common form of distortion is a *radial lens distortion* of image points. During the projection of a point through a typical camera lens or lens system, the projected point gets more displaced the farther it is from the center point in the image, where the optical axis intersects with the camera sensor. A common way to model this type of radial distortions is by means of an infinite series:

$$\mathit{\Delta}\boldsymbol{p}_\mathrm{s} \;=\; \boldsymbol{p}_\mathrm{s}\,(\kappa_1\,\|\boldsymbol{p}_\mathrm{s}\|^2 + \kappa_2\,\|\boldsymbol{p}_\mathrm{s}\|^4 + \ldots) \quad . \tag{2.8}$$

Here $\|\boldsymbol{p}_\mathrm{s}\| = \sqrt{p_\mathrm{sx}^2 + p_\mathrm{sy}^2}$ describes the distance of the undistorted point from the center of projection on the camera sensor and $\mathit{\Delta}\boldsymbol{p}_\mathrm{s}$ stands for the displacement of the projected point due to radial lens distortion, with its distorted coordinates at $\boldsymbol{p}_\mathrm{s} + \mathit{\Delta}\boldsymbol{p}_\mathrm{s}$. Common lenses can be modeled with a maximum of two distortion coefficients $\kappa_1, \kappa_2 \in \mathbb{R}$. Higher order terms might only lead to instabilities during calibration [Tsa87, Hei97]. Although in most cases it is sufficient to model radial distortion to achieve a proper modeling of the real optics, sometimes other types of lens distortions need to be considered as well. For instance, [Hei97] details further tangential distortions that occur due to non-collinearity of the centers of curvature of lens surfaces.

2.5.3 System Calibration and Registration Techniques

In some AR systems, the estimation of initial registration parameters is done in a non-systematic fashion, whereas the updates of their changes is handled by the tracking system. By means of manual interaction, the user establishes a set of correct viewing parameters for the virtual camera and registration parameters for the mapping between virtual and real objects. For instance, the user might be asked to adjust the location of a virtual object while looking at the real counterpart through the AR system from different viewpoints. This interaction has to go on until the registration "looks correct." Usually this type of approach does not lead to very satisfactory results, especially if the user has little experience with that system. A registration might "look right" from one point of view, but not from another. It might also be possible to directly measure some of the registration parameters or derive them from hardware specifications, such as measuring the interpupillary distance with optometrical tools or taking the focal length of a camera lens from its specifications sheet. Nevertheless, these methods usually just lead to limited success and further adjustments become necessary [Azu97].

Common methods for the calibration of AR systems involve view-based tasks. The system asks the user to perform a set of tasks interactively that build geometric constraints. All the input is collected and once the system has enough information it determines the viewing parameters and any other registration parameters. Especially optical see-through methods base their display–user calibration on these kinds of approaches, since the system does not have direct access to the real images that the user sees. Due to the difference in viewing parameters for each user—and even for each new session—this calibration has to be repeated whenever the system gets used. Several methods of this view-based task type have been proposed, such as the single point active alignment method (SPAAM) for the calibration of optical see-through HMDs in [Tuc02], which also lists further references for this kind of interactive HMD calibration. Here, the user aligns image points with a single 3D point in the world coordinate system from various viewpoints for calibration. This method has several advantages to previous calibration methods, as it only requires the user to align a single point at a time and as such reduces the causes of error by making the data collection task simple. The accumulated 3D-2D point references, together with the relative tracking information of the head poses, permit a calibration of the intrinsic camera parameters of the eye/display combination. The technique can be applied for monoscopic and stereoscopic optical see-through HMD calibration.

Collecting point references to calibrate a real camera is a known task in computer vision and photogrammetry. The 2D coordinates of the image points can be extracted from the camera images, and no manual interaction from the user is necessary. The images are available for processing in the computer, unlike the images of an optical see-through AR system, which are formed on the retina of the user's eye. As such, video see-through AR systems have the advantage of objective, user-independent calibration by extracting the 2D points from the camera images. In combination with the corresponding 3D point coordinates, established camera calibration schemes can be employed to estimate the unknown camera parameters. Camera calibration has been studied extensively in the computer vision community [Tsa87, Len88, Wen92, May92, Hei97, Zha00]. Commonly, the unknown parameters of the camera are estimated from point correspondences. As such, AR systems can utilize the collected 3D-2D point correspondences (either through user-interaction in optical see-through or extraction from the camera images in video see-through systems) to calibrate the viewing parameters of the eye/display combination and of the real camera in an optical and in a video see-through system, respectively.

Besides providing proper viewing parameters for the virtual camera, a registration of the virtual to the real objects has to be performed to achieve an accurate AR overlay, as outlined in Chapter 2.5.1. Especially in the context of image guidance, medical imaging information from within the patient needs to be registered properly to the visible surface features of the patient. As Chapter 2.7 will describe and survey existing prototype AR systems and approaches for interventional image guidance, it will also highlight the system-specific approach of performing this alignment of medical dataset and patient if it has been reported in the cited literature. Most approaches can be classified into fiducial-based and surface-based methods. Fiducial-based methods, on the one hand, attach either markers to the skin of the patient, implant them into the bone (such as the skull), or attach them to a dental holder, which the patient securely places between his teeth. The 3D world coordinates of those markers can be measured with a tracked 3D pointer. They are also made of materials that make them easily visible in the medical images. Therefore, the coordinates of the markers in the volumetric medical dataset can be related to the 3D world coordinates, which permits an estimation of a proper coordinate transformation between real and virtual worlds. Surface-based methods, on the other hand, extract the skin surface of the patient from the medical dataset. For the intervention, the surface of the patient is reconstructed by means of laser scanning or structured light and matched with the extracted surface from the dataset providing a coordinate transformation between real and virtual worlds.

2.5.4 Static and Dynamic Errors

One of the main goals during the design of an AR system is to minimize the possibility of errors in the registration parameters. The sources of these registration errors can be classified into two groups: static and dynamic [Azu97]. *Static errors* cause misregistration even when nothing, neither the viewpoint nor the real objects, change over time. *Dynamic errors* only affect the registration parameters during motion, i.e., when either the objects in the environment move or the viewpoint is changing.

Four major sources of static errors can be specified:

- *Mechanical misalignments and optical distortions:* Discrepancies between hardware specifications and the actual properties of the real system, such as orientations and distances between beam-splitters, optics, and monitors in an optical see-through HMD cause registration errors. Subtle changes in position and orientation in the optics can cause significant errors in the fused images that are sometimes even hard to compensate for by renewed calibration and registration. Other distortions, such as optical distortions in lenses of video cameras and display optics, can especially be apparent in AR systems with wide field-of-views, since radial distortion increases with the distance from the optical axis. Straight lines appear curved. Although optical distortions can be modeled and compensated for, the extra mapping step of the computer graphics might introduce more delay in the output of the augmented reality image.

- *Incorrect mapping of virtual to real objects:* Proper overlay of graphics on the real world view is dependent on an accurate mapping of geometric features of the real objects to features of the virtual ones. Usually two separate coordinate systems are given for the real world objects and their virtual counterparts. Even if the coordinates of the real object are accurately known in the coordinate system of the viewer, errors in the mapping between the real object and the virtual one will cause misaligned overlays. One task of application specific registration methods in AR is to provide parameters that map corresponding features between real and virtual objects. For instance, the tip of a biopsy needle should precisely correspond to the tip of the graphical representation of a needle. Slight derivations will cause the graphics of the needle to appear alongside the real needle in the AR view (instead of overlaying it), which can render interventional augmented reality guidance useless. Furthermore, problems might arise when real objects deform over time, for instance, caused by motion of organs inside a patient. This will introduce new registration errors into the mapping function between virtual and real objects.

- *Incorrect viewing parameters:* The projection of the virtual world onto a 2-dimensional plane is dependent on internal camera parameters of the virtual camera that renders the virtual scene, such as focal length or the field of view. These internal parameters have to be derived from the view of the real world, for instance the internal camera parameters of the real camera in a video see-through system. Furthermore, the output of the tracker component, which reports changes in the registration parameters, has to be mapped to actual changes of the viewing parameters. For instance, if an AR system with a video see-through HMD employs a head tracker that reports the location and orientation (i.e., the pose) of the head-mounted tracker sensor in 3D space, this pose of the sensor needs to be mapped to the pose of the head-mounted video cameras, for each eye. In this case correct rotation and translation parameters have to be estimated for the mapping function to define a rigid body transformation between the tracker and the camera coordinate systems. Systematic static errors in the graphics overlay occur if the viewing parameters are incorrect. Calibration techniques can be applied to estimate proper viewing parameters. The previous section listed methods and approaches for this registration and calibration task.

- *Errors in the tracking system:* Distortions in the output of the tracking system usually lead to some of the most serious type of static registration errors. They are typically non-systematic and cannot fully be characterized. For instance a tracker that is responsible to track the location and the orientation of a hand-held needle, might get influenced by surrounding electro-magnetic fields from a CRT monitor if its underlying working principle is based on magnetic tracking technologies. Therefore, often it is not possible to measure these errors precisely or to eliminate them. Chapter 2.6 will discuss this topic further.

Besides static errors, dynamic errors are the other cause of misregistration. They occur because of end-to-end system delays, which cause a lag in the visible response of the AR system to an outside action. This latency occurs because of the time each component requires for processing. As an example, a video see-through system needs to capture the live images, read them from the frame buffers, wait for the tracking component to deliver tracking information to update registration parameters, render the virtual graphics on top of the video images, and refresh the display screens to show the final result. Delays of about 100ms or more are very typical for existing systems.

System delays only cause misregistration when motion occurs. It is the change in registration parameters during motion that causes these errors to become apparent, due to the overall processing time and the delayed update of the registration parameters. An AR system that ignores these errors, will exhibit a delayed response of the computer graphics to actions in the real images that it augments. For instance, a head-mounted optical see-through system would render graphics that fits to the user's view from moments before. If the user moves his head the graphics would lag behind and destroy or seriously distort the perception of an augmented scene with this latency between graphics and real world. The following methods exist to reduce dynamic misregistration:

- *Reduce system lag:* The ideal way of reducing or removing dynamic registration errors is to reduce the time the AR system needs to generate the visual output. Faster graphics accelerator cards allow fast rendering of complex virtual scenes. It is important, though, to limit the graphics to the necessary elements. Complex volume rendering of large medical datasets is still time consuming in the context of a real-time AR system with, i.e., an update rate of 30 stereo images per second. It is also important to utilize parallel pipelines within the computer system. For instance should the read-out and transfer of the camera images from the frame buffers to the main memory or graphics memory of a video see-through system not block the graphics engine, which at the same time can draw and generate the virtual view. The reduction of system delay is possible to some extent, its complete avoidance is not. Many hardware related tasks are part of the delay, such as the refresh of the monitor screen with the contents of the graphics memory or the measurement of the head pose with a tracking system.

- *Match temporal streams:* Video see-through AR systems have the possibility to delay the video streams that are taken of the real world. Every incoming video image can be time-stamped and matched with the tracking information that was calculated from measurements with the same time stamp. The graphics can then be rendered with proper registration parameters for this video image. If handled carefully, this method has the capability to reduce any visible lag between graphics and real world completely in a video see-through system and create a very believable fused world view. The only main disadvantage of this technique is its delay of the real world view. This means the user sees real actions delayed the same way the graphics is delayed. Nevertheless in many applications this overall system latency might be the better alternative, due to the fact that a visual-visual conflict is much more prevalent than the visual-kinesthetic and visual-proprioceptive ones (similar to the discussion about AR versus VR in Chapter 2.5.1). Although, some applications might require a very fast response to real actions, under ignoring the virtual graphics. In those cases a delay of the video stream is not recommended.

- *Predict motion:* If it is possible to predict the future location of the real objects in respect to the viewer, or in general to predict the change in registration parameters, then the virtual scene can be rendered with those predicted values. This is a method that can, for instance, be applied in optical see-through, where the real world always is viewed instantaneously. Nevertheless, application-specific methods need to be researched that permit a successful prediction of the motion. For example, many movements of a user's hand can be very random and hard to cover with existing approaches for motion prediction. However, the combination with other sensors, such as inertial sensors, has the potential to make predictions more accurate.

2.6 Tracking

Fusing images from real and virtual environments requires the continuous update of registration parameters that change over time. Augmented reality systems typically invoke a tracking component that adjusts the parameters of the previously discussed registration in real-time, i.e., ideally a minimum of 24 to 30 times a second. Although these parameters could include intrinsic parameters of the calibration, for instance the focal length, or parameters of the 3D registrations for the alignment of the real and the virtual objects due to deformations, as might occur in patients caused by respiratory motion, typically, the tracking component is responsible for estimating the 3D position and orientation (i.e., the pose) of the viewer and hand-held tools with respect to the real environment. As such, tracking systems are an intrinsic element of AR systems and need to be designed or adapted specifically for the particular requirements of the application that the AR systems target.

2.6.1 Pose Estimation

The *pose* of an object is synonymous for its position and orientation relative to a given world coordinate system. Therefore the pose of objects in 3D space exhibits six degrees of freedom, with three translational and three rotational components. The mapping of a point from world to object-centered coordinates can be described by a rigid transformation

$$\boldsymbol{p}_{\mathrm{o}} = \boldsymbol{R}\boldsymbol{p}_{\mathrm{w}} + \boldsymbol{t}\,, \tag{2.9}$$

where $\boldsymbol{p}_{\mathrm{o}}, \boldsymbol{p}_{\mathrm{w}} \in \mathbb{R}^3$ denote the coordinates of the point in the object and world coordinate system, respectively. Rotation is described by the orthonormal matrix $\boldsymbol{R} \in \mathbb{R}^{3\times 3}$ and translation by the vector $\boldsymbol{t} \in \mathbb{R}^3$. Pose estimation is the process of determining the rigid transformation described by $\{\boldsymbol{R}, \boldsymbol{t}\}$.

In augmented reality applications it is of particular interest to estimate the pose of the viewer with respect to its environment. More specifically, the pose of the virtual cameras, for the left and the right augmented view, as detailed in Chapter 2.5.2, needs to be estimated continuously, unless it is a completely stationary system. In most cases, the intrinsic camera parameters, describing focal length, image center, lens distortions, and more (see equations (2.5), (2.6), and (2.8)), stay constant over time. Which means, that only the extrinsic camera parameters, i.e., the pose, (see equation (2.4)) of the two virtual cameras (or one in monoscopic AR systems) need to be updated with the viewpoint changes of the user. For example, the neurosurgeon's current viewpoint needs to be known with respect to the patient when generating augmentations of the patient's head with medical images.

Besides tracking the viewer it is often necessary to track instruments and tools that the user utilizes to interact with the environment. For instance, the pose of a biopsy needle should be known at all times when generating graphics augmentations of the patient and the needle. The tracking system needs to estimate reliably the position and orientation of the instrument with respect to the patient.

Video see-through AR systems offer the possibility to directly estimate the pose of the camera(s) with respect to a reference coordinate system or even to estimate the pose of instruments and tools that are in the field of view of the cameras of the AR system by means of pose estimation techniques known to the computer vision community (e.g., [Qua99]). Pose estimation has been a fundamental topic in complex computer vision problems, such as object recognition, where the location and orientation of objects has to be determined from camera images (e.g., [Nie90, pg. 182]). As such, the video images from the AR system could be utilized to determine the poses of the cameras and the instruments. These types of computer vision pose estimation techniques in unaltered environments are usually computational expensive and require elaborate image processing strategies to extract necessary image information. Therefore, this methodology is rarely used in video see-through AR systems that require high speeds and robust tracking results, unless the environment is artificially adapted to this pose estimation task, for instance by attaching artificial fiducials that can be easily extracted from the video images.

In recent years, systems for surgical navigation have been developed that utilize tracking solutions to estimate the pose of medical instruments during the intervention [Bir00a]. It seems straight forward to apply these tracking systems in interventional AR guidance as well, where the tracking of instruments and tools is a subtask of the overall tracking problem. Furthermore, the application field of virtual reality has been relying on tracking approaches to estimate the

Figure 2.21: Two-camera optical tracking system *Polaris* with passive retroreflective markers. Infrared light is emitted from LEDs around both cameras and reflected by retroreflective spheres that are attached to the tracked object. *(Picture courtesy of Northern Digital Inc., Waterloo, Ontario, Canada)*

viewpoint of the user to be able to provide an experienced immersion into the virtual environment. These real-time tracking systems have been used to track the position of a head-worn display and hand-held tools for interacting with the virtual environment (e.g., [Vog00]). The real-time tracking solutions that have been developed in those two independent application fields—surgical navigation and virtual reality—are based on the same principles of mostly optical or electromagnetic techniques and will be described in the following sections.

2.6.2 Optical Tracking Systems

Optical tracking systems utilize mostly two or three and sometimes even more cameras that are placed in close proximity to the user of the tracking device. The objects that need to be tracked, such as surgical tools in a setup for interventional navigation, are equipped with fiducials that can be easily extracted from the camera images. These can be either passive markers, made of retro-reflective material, that are illuminated by infrared LED arrays that are mounted next to the cameras or the fiducials can be active infrared LEDs that are sequentially pulsed. The 3D position of the fiducials can be determined from the extracted 2D coordinates in the images of the calibrated cameras by means of triangulation.

Among other applications, these systems are most widely used for navigational support

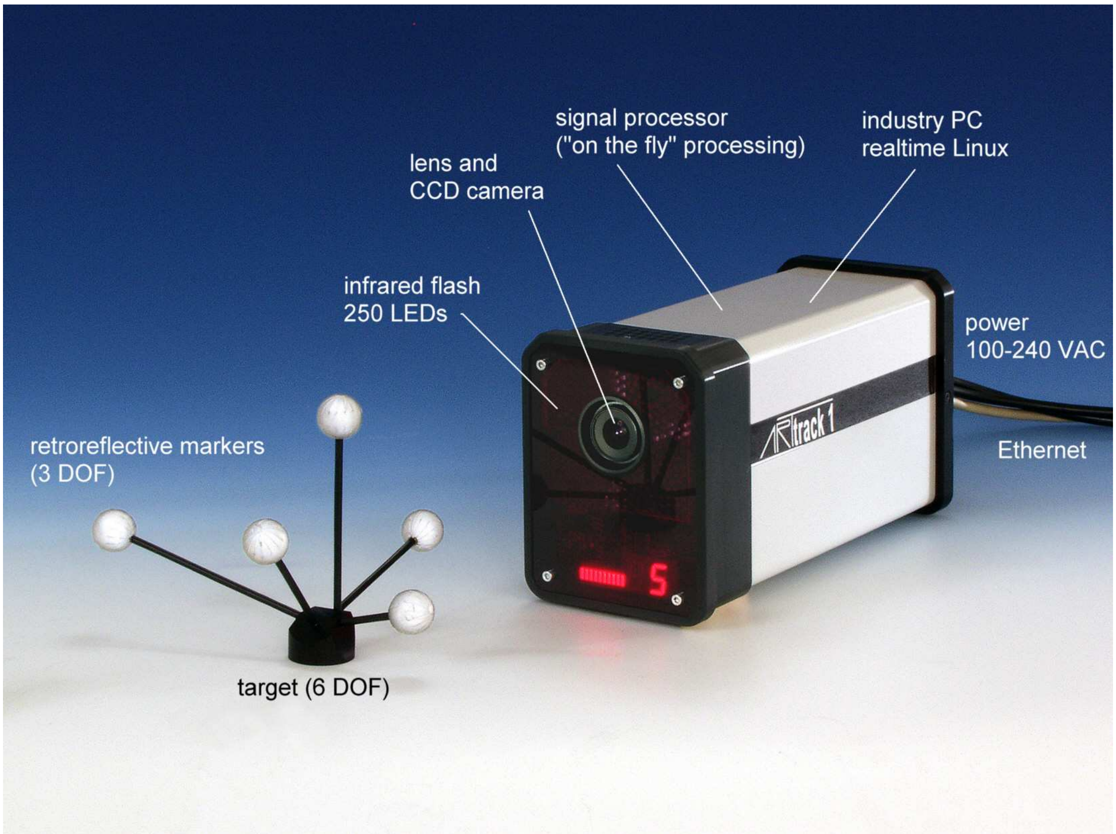


Figure 2.22: Multi-camera optical tracking system (*ARTtrack1*) with passive retroreflective markers, where spatial coverage can be extended by adding more cameras to the system. A minimum of two such cameras must be installed and calibrated to each other before usage of the system. *Targets* are a set of retroreflective spheres, that are assembled in a known spatial constellation and can be attached to the tracked object to track all 6 degrees of freedom. *(Picture courtesy of A.R.T. GmbH, Weilheim i.OB, Germany)*

in computer assisted surgery [Bir00a]. Figure 2.21 and Figure 2.22 show two commercially available systems that are being used in this field. Optical tracking offers the following advantages:

- Unlike mechanical tracking systems, which find applications outside the medical arena, the advantage of optical tracking systems is their separation of the tracked device from the actual tracking system, which is important for a proper sterilization of medical instruments during the intervention.

- Optical tracking provides high precision throughout the working volume independent of electro-magnetic interferences.

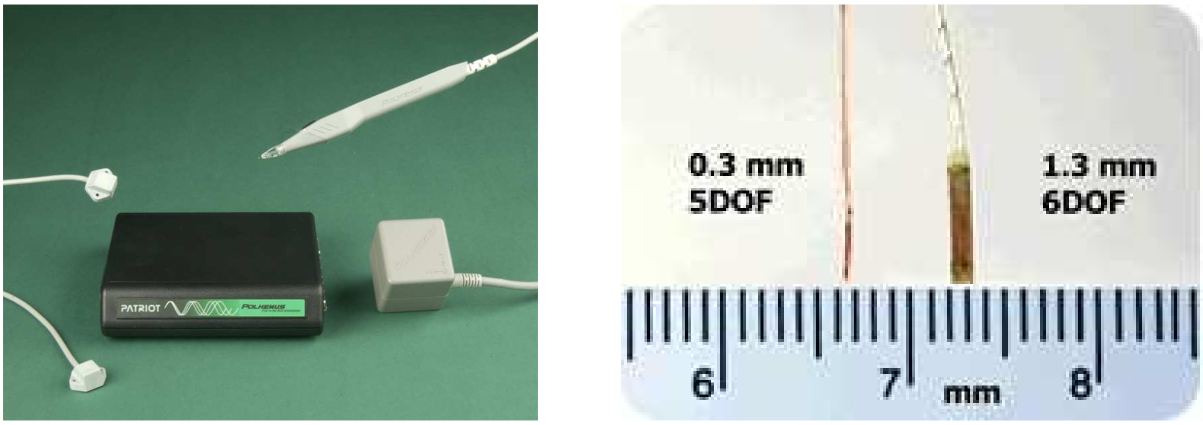


Figure 2.23: Electromagnetic tracking devices. *Left:* AC-based tracking system for sensors with 6 DOF and an update rate of 60 Hz. *(Picture courtesy of Polhemus, Colchester, VT, USA) Right:* Sensors for pulsed DC electromagnetic tracking for use in miniature devices. *(Picture courtesy of Ascension Technology Corporation, Burlington, VT, USA)*

Besides these advantages, optical tracking has the following shortcomings in respect to other tracking solutions:

- The line-of-sight between the fiducials and the cameras has to be guaranteed throughout the usage of the system. The user cannot step between the tracking system and the tracked instruments. This can lead to cumbersome setups when the working area is crowded and the placement of the cameras must be chosen well to guarantee reliable tracking results.
- Optical systems with fast acquisition rates are expensive due to their need for proper computing power and precisely calibrated high-quality cameras.
- For the measurement of the orientation of tracked instruments several markers need to be attached, spanning a certain dimension in space, whose positions are tracked. This can lead to cumbersome devices.

2.6.3 Electromagnetic Tracking Systems

Electromagnetic fields can be utilized for the tracking task. Existing tracking systems based on this approach utilize either alternating (AC) or quasi-stationary direct current (DC) pulsed coils to generate the required field. To determine the position, a sensor assembly is used that consists of search coils (AC trackers) or fluxgate sensors (DC trackers).

Due to its flexibility, electromagnetic tracking has found applications in the medical arena. Figure 2.23 shows an example of an AC tracking system and miniature DC-pulsed sensors. The major advantages of electromagnetic tracking systems are:

- The main shortcoming of optical tracking, an unobstructed line-of-sight, is not present in electromagnetic tracking systems. Since the electromagnetic field penetrates non-conducting materials, the user can easily step between the tracking system and the tracked instruments, which can be very beneficial when the workplace is very restricted.
- Electromagnetic trackers can provide small sensors with very fast update rates (e.g., 120 Hz), while their price can be kept low.

Nevertheless, there are drawbacks of electromagnetic tracking that are of major concern in certain application areas, such as interventional navigation:

- Ferromagnetic and conductive materials in the vicinity of the tracking device cause distortions and decrease the accuracy of the tracking results. Moreover, it is difficult to quantify the amount of error introduced into the system during tracking if equipment with these kinds of materials, e.g., metal objects, CRT screen, etc., are placed in the proximity of the tracker.
- In the context of interventional navigation, sterilization of the sensor assemblies is limited to other methods than the common autoclavation of medical instruments.
- The coils that are tracked have to be attached to cables supplying the current, unlike passive fiducials of optical tracking systems that can be detached from anything else besides the tracked object.

2.6.4 Other Tracking Systems

Mechanical arms with angle sensing potentiometers were used as one of the earliest devices for pose determination. The difficulty in operation, the need for separate arms for each tracked object, as well as basic problems in sterilization make this approach unsuitable for interventional navigation.

Another tracking approach is based on ultrasound spark-gap emitters that are localized by highly sensitive microphones and calculations based on the speed of sound. Similar to optical tracking systems, the line-of-sight restriction applies to ultrasound-based tracking. Moreover, the variation of the speed of sound due to different conditions of air humidity and temperatures decrease the accuracy of these systems considerably.

2.7 Existing Approaches for Interventional AR Guidance

In the realm of interventional image guidance different approaches to augmented reality have been investigated and evaluated in recent years [Sau08]. Although each reported prototype system brings its own very unique features and is targeted toward a certain set of medical applications, they can be classified into a few groups that distinguish their methods of fusing real and virtual worlds and their presentation of the augmented medical scenario to the physician. The following survey describes examples for systems with their corresponding applications that are based on optical microscopes, video see-through technologies, large transparent screens, tomographic overlays, overlays in video endoscope, and other methods. Applications that are targeted by those systems include neurosurgery and otolaryngology, cranio- and maxillofacial surgery, breast and abdominal needle biopsies and tumor ablations, orthopedics, and cardiovascular and thoracic surgery.

2.7.1 Optical Microscope Systems

Operating microscopes are being used for many neuro- and ENT surgical procedures. Interventional image guidance can be achieved by overlaying precisely aligned 3D graphics, which is derived from the patient's preoperative images, into the optics of the microscope. Proof of principle and early phantom tests have been presented, e.g., in [Fri89] and [Edw95a, Edw95b].

Based on this approach, a prototype augmented reality system has been developed at Guy's Hospital, London, that provides 3D surgical navigation for microscope-assisted interventions in neurosurgery and otolaryngology [Edw00, Kin00, Edw99]. *MAGI* (Microscope-assisted guided interventions) is based on a Leica binocular operating microscope with an external mount. Medical information, which is derived from MRI or CT scans of the patient's head or neck, is integrated as overlays into each eyepiece of the stereoscopic microscope. Two dedicated monochrome VGA (640×480) displays incorporate the computer-generated images via beam-splitters into the left and the right microscope view. Figure 2.24 shows a picture of the system in use and the augmented view through one of the oculars of the microscope.

The spatial position of the microscope is tracked with an optical tracking system (Optotrak, Northern Digital Inc.) and a set of infrared LEDs that are attached to the microscope. To achieve highest precision for the registration of the medical images to a patient-aligned coordinate system, bone-implanted markers are used, which are inserted prior to preoperative imaging. The coordinates of those fiducials are extracted in the medical images and can be marked in patient space with a tracked pointer.

Figure 2.24: *Left:* A device for microscope-assisted guided interventions (MAGI), developed at Guy's Hospital, London. *Right:* Augmented view through the microscope. *(Pictures courtesy of Philip Edwards)*

To be able to relate the coordinate systems of the patient and the microscope to each other either the patient's skull needs to be completely stabilized or the movement of the skull must be tracked during the procedure. To avoid the necessity of a neurosurgical head clamp, which fixes the patient's head to the operating table and provides a stable patient coordinate system, a locking acrylic dental stent (LADS) has been designed which attaches to the patient's upper teeth and contains a set of infrared LEDs that can be tracked with the Optotrak system. Furthermore it is being evaluated if accuracy permits to avoid bone-implanted markers and to replace them by a set of MRI/CT-fiducials that are attached to the LADS device instead.

As has been shown in several clinical tests [Edw99], an operating microscope with augmented reality image guidance can be used to perform difficult surgical interventions on head and neck. For instance, the removal of a petrous apex cyst, where a more anterior approach was necessary to preserve hearing, was successfully guided with the MAGI device. The precision of the graphics overlay was reported to be below 1mm throughout the procedure.

A similar approach to MAGI has been taken at the AKH Vienna, Austria, (later the CARCAS-Group at the University Hospital Basel, Switzerland) to provide interventional image guidance by means of augmented reality, where medical information is injected into the optical path of a miniature head-mounted operating binocular. Clinical applications of operating binoculars, which typically have a 3–7$\times$ magnification, include oral and cranio-maxillofacial surgery, plastic and reconstructive surgery, as well as orthopedics. The commercially available *Varioscope*, a head-mounted, lightweight operating binocular (Life Optics, Vienna), with autofocus, automatic parallax correction, and zoom, has been modified for AR visualization of medical imaging information [Bir02, Bir00c, Bir00b, Fig01].

Figure 2.25: *Left:* The Varioscope AR, a head-mounted operating binocular which has been modified to an augmented reality device for interventional image guidance, developed at AKH Vienna. *Right:* Augmented view through one of the oculars, for the scenario of endosteal implant insertion. *(Left picture [Bir02] ©2002 IEEE; Pictures courtesy of Wolfgang Birkfellner)*

Figure 2.25 shows a picture of this prototype device, also referred to as the *Varioscope AR*, and the augmented view through one of the oculars. As several other optical instruments, the original Varioscope contains prisms for image rectification (to correct for an inverted image) in both optical paths. To realize the Varioscope AR two miniature LCD displays with VGA (640×480) resolution are attached to those image rectification prisms of the Varioscope. The prisms, modified on one side with a thin semi-transparent layer to inject the LCD image into the optical path, thus act as beam splitters. The ocular magnifies the image from the main lens as well as the image from the projection optics. Since both images are provided in the same focal plane of the main lens, this approach solves the problem of realizing a commonly focused view for virtual and real scene. Figure 2.26 illustrates this unique principle of merging real and virtual images by an optical system. Many details have to be taken into consideration to achieve a precise calibration of such a specific optical AR system with variable zoom and focus [Fig05, Bir01].

The Varioscope AR was designed to work with existing CAS systems. To study this AR device it was integrated into a surgical navigation system, also developed at AKH Vienna mainly for cranio- and maxillofacial surgery, which employs optical tracking. The tracking device of the CAS system, which could be either the Flashpoint 5000 (Image Guided Technologies) or the Polaris (Northern Digital Inc.), was used to track the position of the HMD. Technical details of the control unit that connects the HMD to the CAS system can be found in [Fig02a, Fig02b].

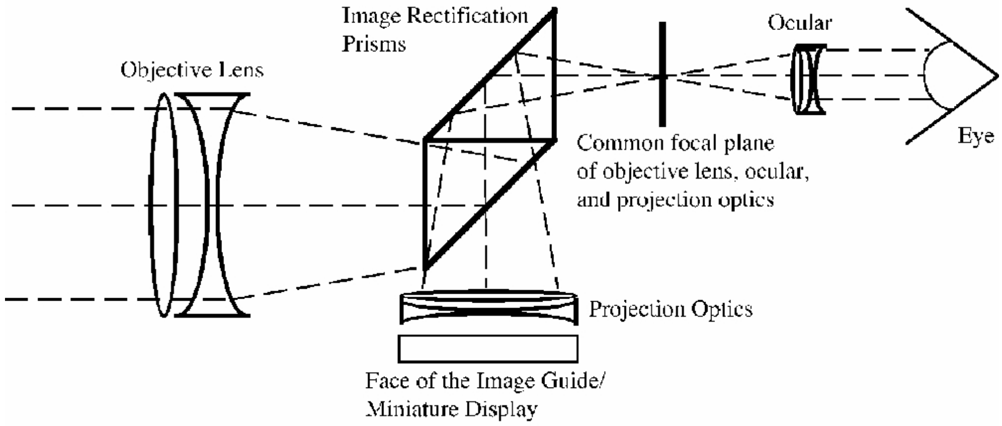


Figure 2.26: The image rectification prisms of the Varioscope are exploited to inject the virtual image into the optical path, which provides a common augmented image in same focal plane of the main lens. *(Picture [Bir02] ©2002 IEEE; Picture courtesy of Wolfgang Birkfellner)*

2.7.2 Video See-Through Systems

Augmented Reality systems that incorporate video cameras to capture the view of the real world, augment this video view with computer graphics in real-time and then display the fused view for the user on a screen. They are commonly referred to as *video see-through* systems. A system, which is based on this approach, was developed at the MIT AI Lab for the purpose of image-guided neurosurgery and studied in close collaboration with the Surgical Planning Laboratory at Brigham & Womens Hospital [Gri98, Gri95, Gri99].

A camera that is placed close to the surgical scene provides the live video feed of the patient. The patient's head as well as the surgical instruments are tracked by an optical tracking system (Flashpoint, Image Guided Technologies Inc.) by means of LEDs that are attached to the surgical instruments and the neurosurgical head clamp, which stabilizes the patient's head. To register the medical information from the MRI scan to the actual patient position in the head clamp, 3D surface points of the patient's scalp are collected with a laser scanner or a tracked pointer. The collected points are precisely registered with the skin surface, which is extracted from the patient's MRI scan [Gri96, Gri94]. During the interventional procedure these registration parameters are used to align medical information from the MRI scan with the video image of the patient's head. The augmented video image on the monitor screen displays the patient in a transparent fashion, with internal anatomical structures from the MRI dataset overlaid on the video of the skull. The tracked pointer is visible as well in this augmented view. Besides the augmented view, the monitor displays three orthogonal MRI slices in separate windows, according

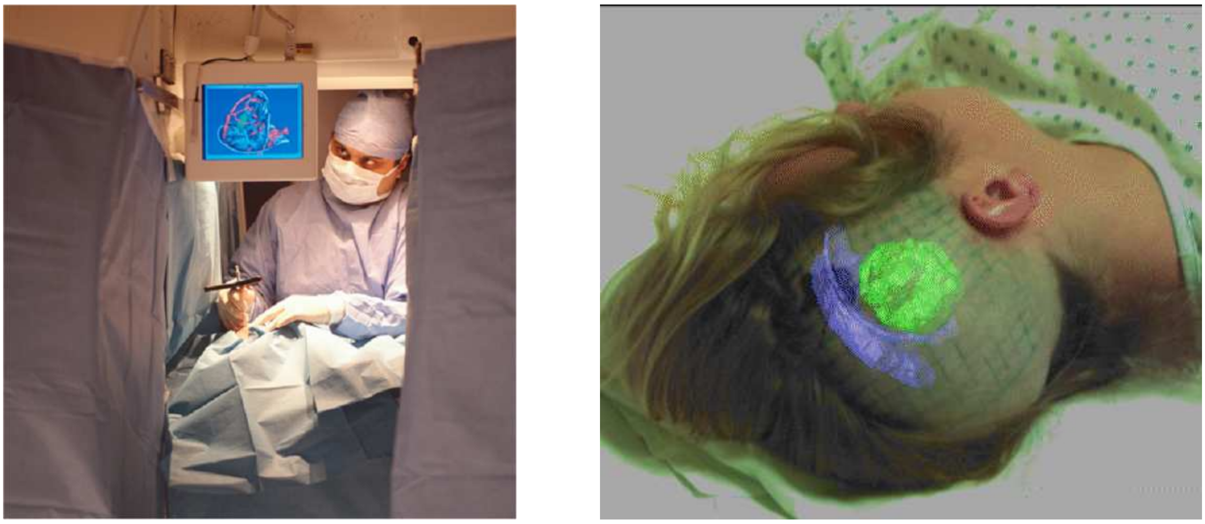


Figure 2.27: *Left:* Monitor-based navigation system for image-guided neurosurgery. *Right:* Patient's head augmented with internal structures, which were extracted from an MRI scan. *(Picture [Gri96] ©1996 IEEE; Picture courtesy of W. Eric L. Grimson)*

to the 3D pointer coordinates, similar to more traditional navigation systems. Figure 2.27 shows a picture of a neurosurgical intervention with the guiding monitor screen above the surgical site and an example image of an augmented view of the patient.

In [Gri98] it is reported that this image-guided neurosurgery system has been used on 70 patients. It effectively supported the surgery in planning the bone cap removal, identifying margins of tumor, localizing key blood vessels, and orienting the surgeon. A wide range of neurosurgical cases were selected to evaluate the efficiency of the system, including tumor resection, pediatric epilepsy, meningioma, and biopsy cases.

AR visualization in the medical field has first been suggested and investigated at UNC, Chapel Hill, for ultrasound-guided procedures [Baj92, Sta94]. A stereoscopic head-mounted display (HMD), which is equipped with two miniature video cameras and a tracking device, is the centerpiece of their video-see-through system. A tracked HMD that displays virtual and real world fused together provides stereoscopic and kinetic depth cues while the user is able to move around and look at the augmented scene from different viewpoints.

Early developments of the UNC system were targeted toward visualization of ultrasound images within the womb of a pregnant woman (left in Figure 2.28). A 3D representation of the fetus could be seen in its actual location. Further research adapted this video see-through approach for ultrasound guided needle biopsies [Sta96, Fuc96, Ros01, Ros02, Sta03]. An adaptation of an earlier version of the UNC system for laparoscopic surgery is described in [Fuc98].

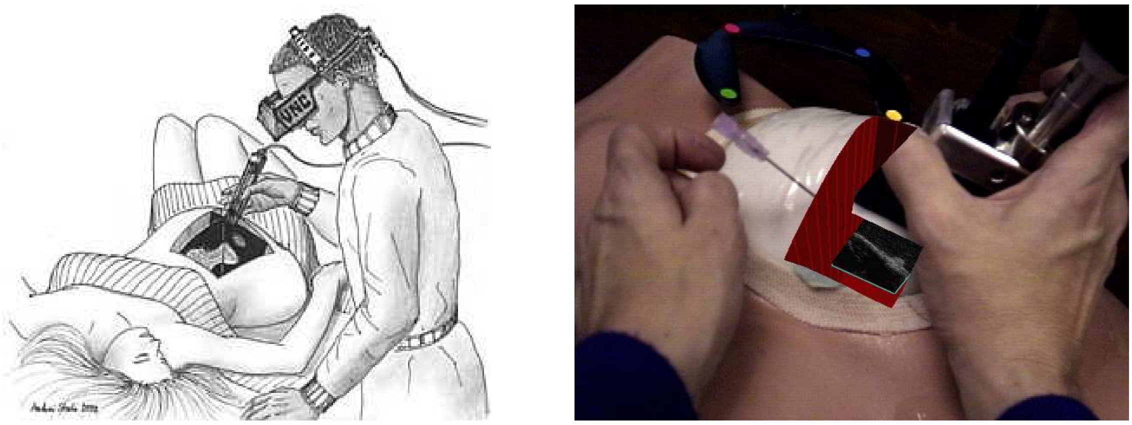


Figure 2.28: Approach for head-mounted video see-through augmented reality at UNC, which provides stereoscopic and kinetic depth cues of the fused scene. *Left:* Schematic representation of an in-situ fetus visualization. *Right:* View through the HMD during ultrasound-guided needle biopsies on breast phantom. *(Right picture [Sta96] ©1996 ACM, Inc. reprinted by permission; Left artwork and right picture courtesy of Andrei State)*

The system, reported in [Ros01], is tracked by an optical tracking device (FlashPoint 5000, Image Guided Technologies), which also tracks the ultrasound probe and a biopsy needle. The ultrasound image is visualized in its actual location, within the patient. Virtual objects identify the location of a breast tumor in the augmented view of the patient. The needle is guided to its target with in-situ ultrasound visualization and augmenting computer graphics (right in Figure 2.28). A randomized, controlled trial to compare standard ultrasound-guided needle biopsies to biopsies performed with the AR prototype guidance system at UNC was used for further evaluation of the system. Fifty biopsies of breast phantoms, performed by a radiologist, were precisely evaluated and it was found that the method with head-mounted AR visualization resulted in a significantly smaller mean deviation from the desired target than the standard ultrasound-guided method on a monitor screen (2.48mm for standard versus 1.62mm for AR).

The development of dedicated stereoscopic video see-through HMDs that are tracked in 3D space poses many technological challenges. In the medical field, video-see-through AR systems with a stationary monitor and camera setup, similar to the one developed at MIT AI Labs, can be seen as the closest relatives to classical image guidance systems, where a dedicated monitor screen in the OR visualizes the tracked probe or instruments virtually within the medical dataset of the patient. Noteworthy is an early prototype of a monitor-based video see-through augmented reality system for neurosurgical procedures that was described in [Lor93, Lor94]. Analog blending of the video output of a computer graphics workstation, displaying the patient's dataset, with live video of the patient lead to an augmented or enhanced reality perception. Also currently,

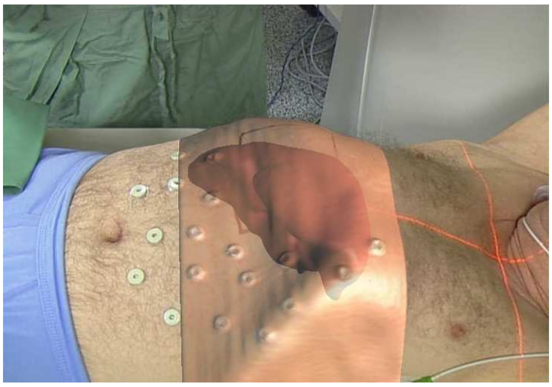


Figure 2.29: *Left:* A data-fusion display system that combines video-view of the patient with intraoperatively acquired scans of a mobile x-ray C-arm in an augmented reality fashion, developed at Jikei Univ. School of Medicine, Tokyo, Japan. *Right:* Augmented reality guidance for liver punctures with a monitor-based video see-through prototype system developed at IRCAD University Hospital, Strasbourg, France. *(Left picture [Hay05] and right picture [Nic05a] with kind permission of Springer Science and Business Media)*

the concept of a single stationary monitor for displaying the augmented video-view of the patient is still of great interest and is being pursued by different research groups. For instance, [Hay05] describes a prototype system, which has been developed at Jikei Univ. School of Medicine, Tokyo, Japan, where intraoperatively acquired CT-like images from a mobile C-arm x-ray system are used to overlay the patient's internal anatomy onto the camera view of the patient in the operating room (left of Figure 2.29).

A monitor-based video see-through prototype system to guide liver punctures for radio-frequency tumor treatment is being developed at IRCAD University Hospital, Strasbourg, in collaboration with INRIA Sophia-Antipolis, France [Nic05b]. In a stationary abdominal phantom the system achieves a target precision of 3mm. First in vivo experiments have been presented in [Nic05a]. The right side of Figure 2.29 shows an augmented view of a patient. The common problem of respiratory motion for interventional guidance of abdominal procedures remains and currently restricts the application of this system to large targets, with a diameter above 3cm.

A unique system for needle placement procedures under video-augmented x-ray guidance with a mobile C-arm has been reported in [Nav99a, Mit00]. The optical axes of both, x-ray and video camera, are aligned by a double mirror system as shown in Figure 2.30. Therefore, x-ray and video camera images of the patient are taken from the same viewpoint. The video image shows the surface of the patient's body and objects located in front of it, while the x-ray image shows the inside of the body. With this approach real-time augmentation of both images can be achieved to support needle placement procedures under AR guidance instead of fluoroscopy.

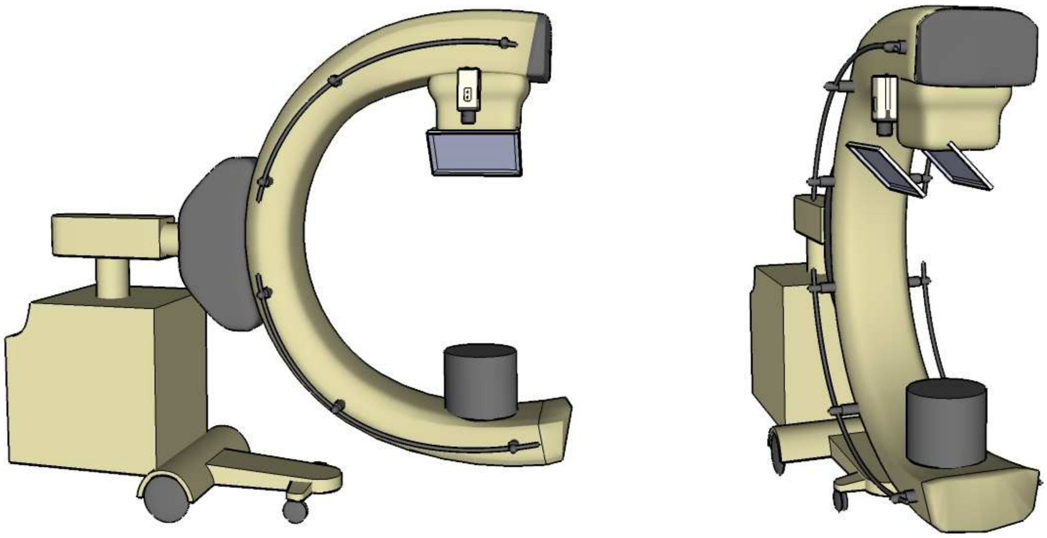


Figure 2.30: Schematic drawing of CAMC. A mobile C-arm is equipped with a video camera and a double mirror system, aligning the video view with the x-ray view. *(Pictures courtesy of Joerg Traub)*

The actual alignment of the needle to the target structure is guided by video images that are co-registered with the x-ray image. This method has the potential to reduce x-ray exposure to patient and surgeon.

2.7.3 Large Screens

Besides the previously introduced methods of augmenting the view of the surgical site with medical information by means of either injecting graphics into the optical path of a surgical microscope or adding graphics to the video view of the patient within the computer, a third approach has been investigated that utilizes a single large screen in form of a semi-transparent mirror. At the CMU Robotics Institute, the MRCAS group (Medical Robotics and Computer Assisted Surgery) developed an image overlay system based on a semi-transparent mirror which is placed above the surgical workplace [Bla00, Bla98b, Bla98a]. A high resolution flat LCD panel is mounted above the half-silvered mirror or beam splitter glass. The physician looks at the patient through the glass while seeing a reflection of the LCD display within the patient at the same time. This constellation creates the illusion of seeing the image below the glass virtually inside the surgical workplace. The display/mirror system is attached to an articulated arm. An optical tracking system (OptoTrak) tracks patient, the display, and the viewer by means of attached LEDs. Figure 2.31 illustrates the concept and its realization as a prototype sys-

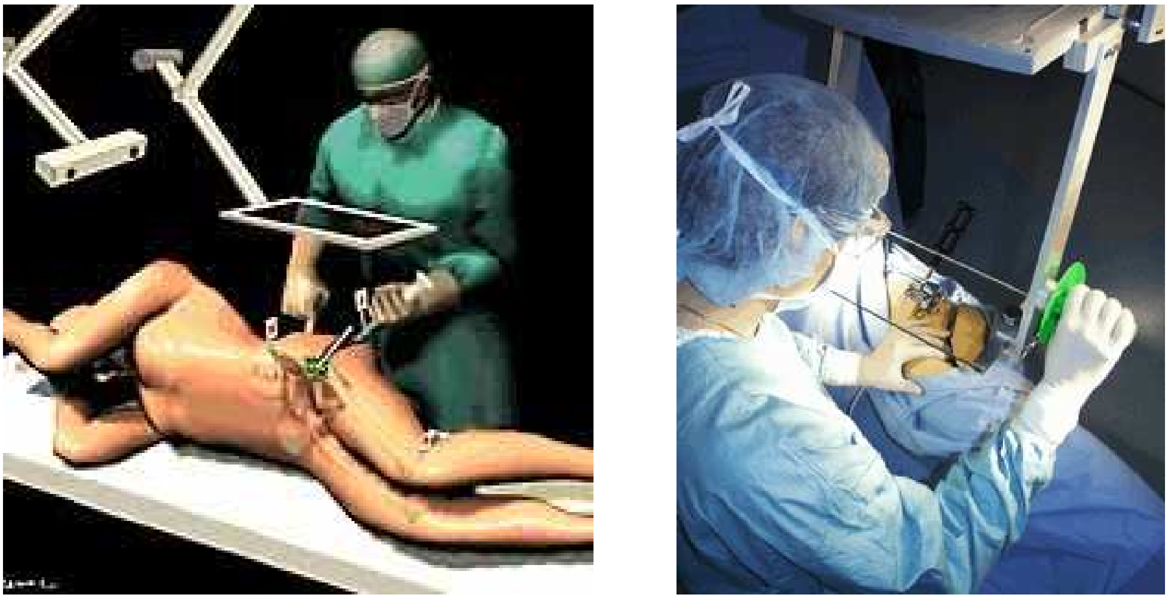


Figure 2.31: Large screen image overlay system for interventional image guidance or surgical education, developed at the Robotics Institute at CMU. *Left:* Illustration of the concept. *Right:* Prototype system in use. *(Pictures courtesy of the Carnegie Mellon Robotics Institute)*

tem. Potential applications include orthopedic surgery, neurosurgical procedures, and surgical education [Bla98b].

In [Bla95] an earlier prototype system from CMU, based on a CRT screen, was described, which provided proof of concept of the display/mirror approach. Additionally, to create a 3D virtual image, shutter glasses were used while the screen renders a different view for each eye, synchronized to the shutter. This provides a stereoscopic depth cue for the virtual scene. Other types of approaches have to be considered to achieve stereoscopic images for an LCD type of display, such as the use of polarized glasses.

Figure 2.32 shows another image overlay system that follows the transparent screen approach [Goe03]. The *ARSyS-Tricorder* has been developed by multiple institutions in collaboration with Fraunhofer-Gesellschaft and utilizes a setup with a stereoscopic projector instead of a display to generate the graphics. The graphics is reflected by the half-transparent mirror above the patient. The physician wears polarized glasses, which give him stereoscopic vision of the virtual images through the mirror. The physician, as well as the mirror and the projector are tracked to provide proper registration between the medical images and the patient.

An image overlay system that overcomes the limitations of the 2D representation of medical images without a stereoscopic approach is described in [Mas00]. This system, referred to as the *slice-display*, is based on the known approach with a flat screen, 2D cross-sectional medical

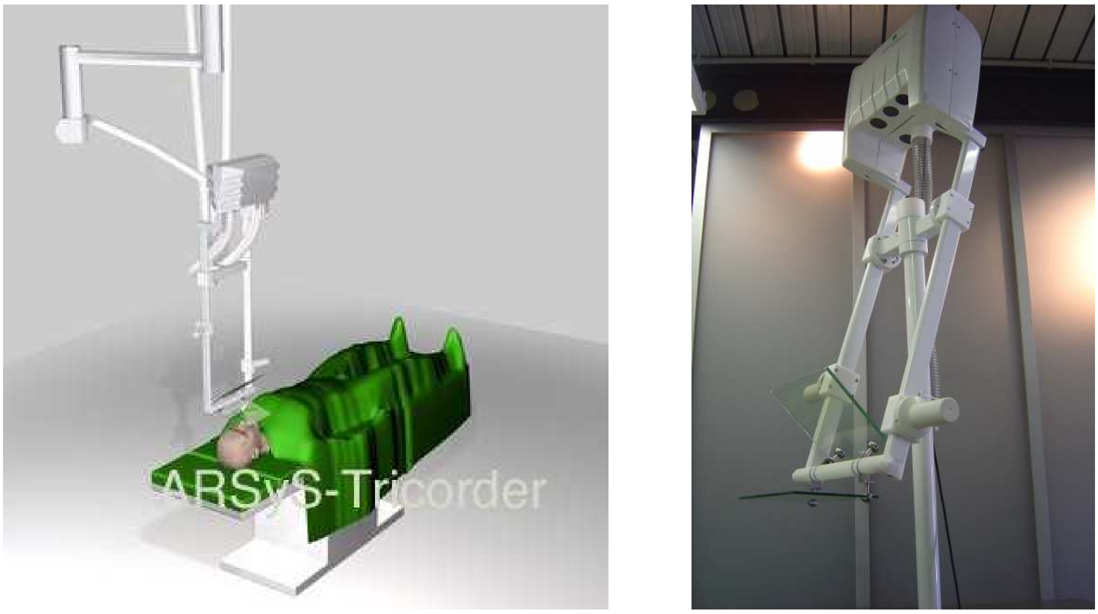


Figure 2.32: ARSyS-Tricorder—Image overlay system with a large screen for interventional image guidance. *Left:* Illustration of the concept. *Right:* ARSyS-Tricorder prototype system. *(Pictures ©Fraunhofer Institut Intelligente Analyse- und Informationssyteme (IAIS), Sankt Augustin, Germany)*

images, and a half-silvered mirror, but the screen has one degree of freedom and can be moved along the line-of-sight (i.e., perpendicular to the screen surface) by the user. Considering that the virtual image of the screen appears within the patient under the mirror, this virtual image "slices" through different parts of the patient's anatomy. A dedicated registration method of the pre-acquired volumetric medical data and the patient below the semi-transparent mirror registers the volumetric patient data in a way that permits moving the screen along its one degree of freedom. This results in slicing the 3D data in different positions, where the virtual reflection of the chosen slice appears to be inside the patient spatially at its correct anatomical location.

At the University of Tokyo another approach of achieving three-dimensional virtual overlays by means of integral videography is being pursued, which does not employ any kind of goggles [Lia04, Lia01, Nak00] but an autostereoscopic display. This system overlays 3D medical images onto the real patient by means of a half-silvered mirror above the surgical site. The specialty of this system is its display, consisting of a high-resolution LCD with a micro convex lens array. It is based on the principle of integral photography, or *integral videography* in this case, which can display geometrically accurate 3D images and reproduce motion parallax without using special glasses or tracking devices. Figure 2.33 illustrates the principle behind this approach and shows a prototype system.

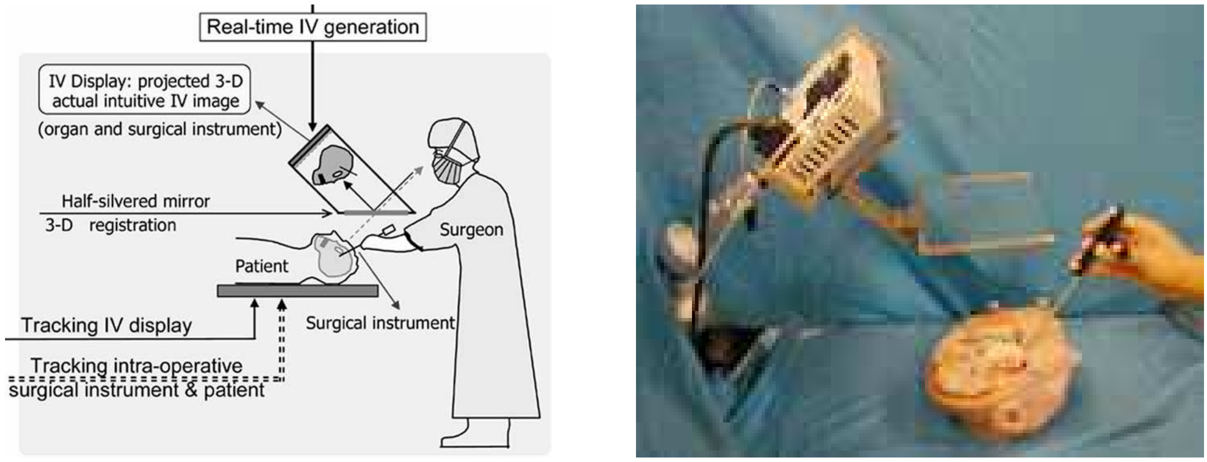


Figure 2.33: Surgical navigation by autostereoscopic image overlay of integral videography, developed at University of Tokyo. *Left:* Illustration of the concept, where the surgeon perceives 3D image overlay without the need for any kind of stereoscopic goggles. *Right:* Prototype system. *(Picture [Lia04] ©2002 IEEE)*

An optical tracking system (Polaris, Northern Digital Inc.) keeps track of the positions of surgical instruments and the target during the surgery. Both, surgical instruments and 3D patient data, are represented and overlaid as IV images during the surgery. First phantom experiments show that needle placement procedures could be guided with this system, with a mean error below 3mm [Lia04]. The pixel density of the display and the lens pitch are the main contributing factors for the quality of the IV image. IV image resolution remains an issue due to the technical challenges of lens-array and screen resolution. Approaches to build higher resolution IV displays are being investigated, e.g., in [Lia02].

An earlier approach to augment a neurosurgical site with 3D images, involving integral photography, was reported in [Mas95, Mas96, Ise97], where stationary integral photography images of three-dimensional medical data were recorded on film. During surgery those integral photographs were superimposed on the patient with a half-silvered mirror. Since this *Volume-graph* does not involve computer screens, but conventional film, the image could not be altered during the surgery. The recording of those images on film took several hours.

2.7.4 Tomographic Overlay

The concept of a half-silvered mirror to augment the patient with medical information from pre-operative scans during a surgical procedure can be directly applied to an interventional imaging device in a way that permits real-time in situ visualization of tomographic images. At the

Figure 2.34: A prototype of the *sonic flashlight*, developed at VIA Lab, Pittsburgh. *Left:* A small flat-panel monitor is attached to the handle of an ultrasound probe. A half-silvered mirror bisects the angle between ultrasound slice and display to create an in situ virtual image. *Right:* Example of a tomographic overlay with the sonic flashlight. *(Pictures courtesy of George Stetten)*

Visualization and Image Analysis (VIA) Laboratory (based at the University of Pittsburgh and Carnegie Mellon University, Pittsburgh, PA, USA) the term *Real Time Tomographic Reflection (RTTR)* has been coined for this approach, which merges the visual outer surface of the patient with a simultaneous scan of the patient's internal anatomy by means of a half-silvered mirror (US Patent 6,599,247) [Ste00, Ste01b, Ste01c].

Figure 2.34 shows a prototype of the introduced *sonic flashlight* at VIA Lab, which augments direct human vision with real-time ultrasound images. A panel type display attached to a B-mode ultrasound probe displays the live ultrasound image. This image is reflected in a half-silvered mirror, which is attached to the probe in a way that the perceived virtual image of the ultrasound scan appears to be in the location of the actual place where the scan is originating from. It provides an illusion of a 2D cross-section through a 3D object, without the need of a tracking device or virtual-reality glasses. Figure 2.35 illustrates this principle.

Subsequent research has adapted this idea to, e.g., magnified real-time reflection of ultrasound for remote procedures [Ste01a], a C-mode sonic flashlight for a matrix array ultrasound probe [Ste03, Ste05], and integrated the sonic flashlight with a laser guide for needle procedures [Wan05]. In [Cha05], the sonic flashlight has been used for a cadaver study in a neurosurgical context where a needle was guided toward a lesion in the brain. Compared to conventional

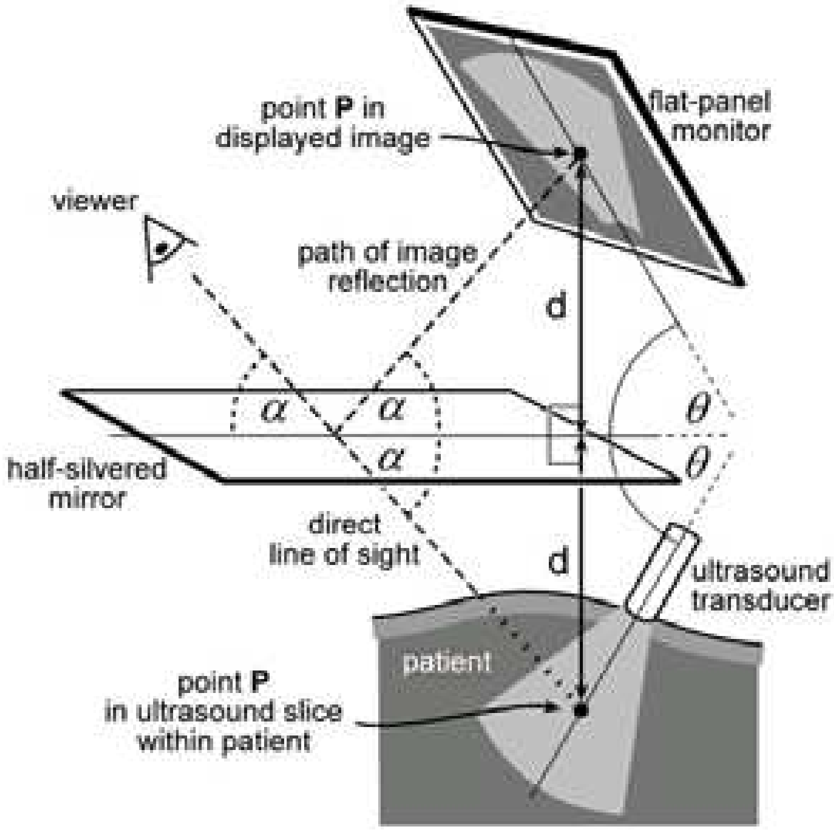


Figure 2.35: The principle of real-time tomographic reflection as implemented in the sonic flashlight. Due to the geometric relation between half-silvered mirror, flat-panel display and ultrasound transducer, each point in the virtual ultrasound image is precisely located at its corresponding physical 3D location. *(Picture courtesy of George Stetten)*

ultrasound procedures with a separate monitor screen and displaced hand-eye coordination, it is reported that the normal hand-eye coordination with in-situ ultrasound of a sonic flashlight helped to guide the needle easily and intuitively into the lesion.

The tomographic overlay concept, where no tracking is needed and which has been realized in form of the sonic flashlight for ultrasound, can be adapted to other imaging modalities. At the CISST Lab at Johns Hopkins University, a prototype system has been developed based on this principle to support percutaneous therapy performed inside a CT scanner [Mas02, Fic04, Fic05a, Fic05b]. A slice image overlay system, which comprises a flat LCD display and a half-silvered mirror, has been attached to the gentry of a CT-scanner. Watched through the semi-transparent mirror, the CT scan appears to be inside the patient virtually at the position in the gentry where the scan is being taken.

Figure 2.36: Tomographic overlay systems for percutaneous therapy inside a CT or MRI scanner developed at CISST Lab at Johns Hopkins University. *Left:* The prototype system with monitor and half-silvered mirror attached to the CT gantry during a cadaver needle placement experiment. All observers see the 2D cross-sectional CT image in the same correct position in 3D space. A marker is attached to the needle, indicating the length of the needle that has to be inserted to reach the target at the correct depth. Otherwise, the needle is not tracked, and its actual position inside the patient can only be assessed with a CT control scan. *Right:* The overlay system constructed around an MRI scanner bed guiding a needle placement procedure. *(Pictures courtesy of Gabor Fichtinger & Gregory Fischer)*

Figure 2.36 shows the system on the left side. This prototype system has been evaluated for needle placements in phantoms and cadavers, guided by the tomographic overlay of CT slices in the scanner. Skeletal targets could be reached with one insertion attempt. Liver targets could be assessed successfully, although tissue deformation poses a challenge. The system has the potential to reduce x-ray dose, patient discomfort, as well as procedure time by reducing faulty insertion attempts during CT-guided needle placement procedures.

The right side of Figure 2.36 shows a similar approach to augmented reality from the same group at the CISST Lab at Johns Hopkins University, to guide needle placement procedures on a closed bore MRI scanner [Fis07, Fis06]. A target application is MR arthrography (MRAr), where conventionally a needle is driven under flouoroscopy or CT guidance into a joint and subsequently a diagnostic assessment of injected contrast is made based on MRI images of the contrast-injected joint. Pre-clinical trials of the proposed AR-guided procedure on the MRI scanner bed (replacing the flouoroscopy or CT guidance) resulted in repeatedly successful first-attempt needle placements into the joint of porcine and human cadavers and therefore show the potential of the system to effectively support and simplify the overall arthrography procedure by eliminating radiographic guidance during contrast injection.

2.7.5 Video Endoscope Systems

Endoscopes and laparoscopes are well established tools to support minimally invasive interventions. Some type of augmented reality guidance can be achieved when the pose of the endoscope-like instruments is being tracked and the video images from the inside of the patient's body taken through those instruments are related to and visually fused with the patient's medical images in real-time, such as pre-operatively acquired CT or MRI.

The system in [Sha98], developed at the Image Guidance Laboratory at Stanford, employs a stereotactic endoscope that is tracked by an optical tracking system. Six CT and MR visible markers, which are placed on the patient, help to register the pre-operative CT or MR images to the patient coordinate system. The reported system combines view-aligned volume-rendered images of the CT or MR data with the video-view through the endoscope and visualizes both together on the workstation monitor. The volume rendering of the pre-operative imaging data is performed in a virtual endoscopy fashion, where the model parameters of the virtual endoscope are derived from the real surgical endoscope and its tracking information. Both images—from surgical and virtual endoscopes—can be blended together to achieve an augmented reality view on the monitor screen. The biggest advantage is that the virtual endoscope can make opaque tissue transparent and give the surgeon a look beyond the visible surface captured by the real endoscope. Furthermore it gives additional guidance information of the surrounding tissue and pathology of the surgical endoscope within the patient.

A similar approach for laparoscopic procedures has been described in [De 01]. The prototype system overlays the video images from the laparoscope with virtual graphics that have been extracted from pre-operative CT scans. An optical tracking system keeps track of the position of the laparoscope in respect to the patient. A potential application can be the visualization of the ureter (extracted from a pre-operative CT scan) as virtual graphics in the laparoscopic view of the pelvis during intervention. Locating the ureter is a common problem during standard laparoscopic surgery.

The augmentation of endoscopic images during robot assisted coronary artery bypass graft (CABG) surgery is described in [Mou03, CM04, Mou01]. The proposed methodology, developed at INRIA by the ChIR Medical Robotics Group, has been studied and evaluated with the *da Vinci* surgical system in animal trials. Figure 2.37 shows the setup and the augmented reality visualization of the coronary tree. A model of the coronary tree, as extracted from preoperative angiograms and CT data, is overlaid on the endoscopic images during this minimally invasive surgery to guide the intervention. An initial registration is based on skin markers which can be related to the heart in the pre-operative images and can be measured during the intervention

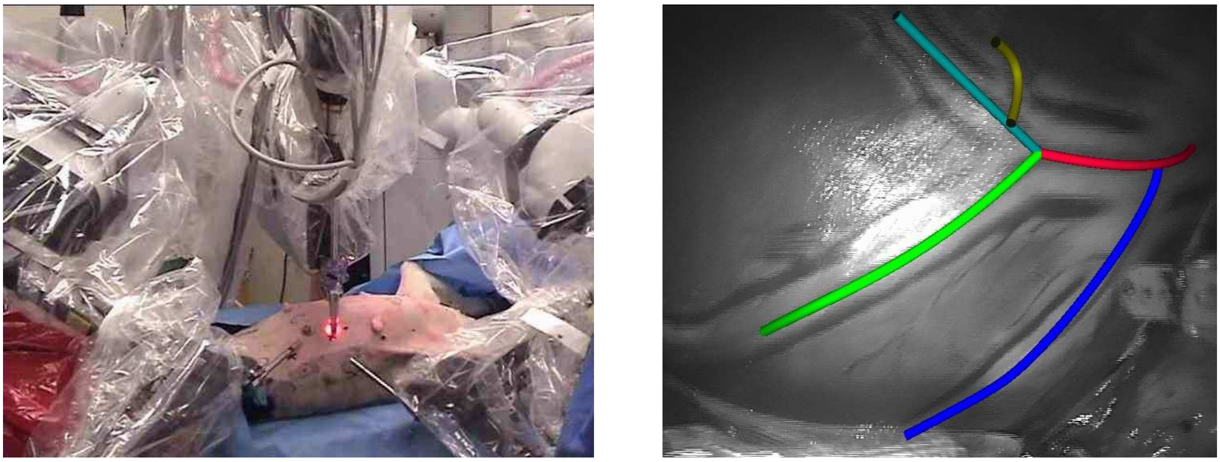


Figure 2.37: Method for augmentation of endoscopic images during robot assisted coronary artery bypass graft surgery with the *da Vinci* system, developed at INRIA. *Left:* Setup of the da Vinci system for an animal study of the proposed method. *Right:* Overlay of a coronary tree on the endoscopic images of the da Vinci system. *(Pictures reproduced with permission from [CM04], ©2004, by permission of SAGE Publications Ltd)*

in respect to the tracked endoscope. A sophisticated interactive method during the procedure allows the surgeon to correct the overlay in real-time. The evaluation of the system on a dog and a sheep shows its effectiveness in helping the surgeon to localize target structures during this robot assisted CABG procedure.

Other approaches for robotically assisted minimally invasive cardiovascular surgery involving endoscopes and augmented reality to merge preoperative planning data with the patient during surgery have been shown, i.e., in [Tra04]. Since the endoscope and other surgical scopes are very commonly used devices for minimally invasive surgery, research topics are plentiful, but mostly do not involve augmented reality approaches. Although some investigated techniques can be classified as mixed reality methods, relating closely to augmented reality, such as in [Dey00], where endoscopic images are merged with 3D surfaces extracted from preoperative imaging data.

2.7.6 Other Methods: Direct Projection

To augment the physician's view graphical information can also directly be projected onto the patient. For practical reasons this direct projection approach is limited to simple 2D graphics that appears on the skin surface, such as points, lines, and contours. The patient is usually draped around the surgical site and does not provide a good projection screen. As a projected image moves and gets distorted when location and shape of the screen change, precise measurement of location and shape of the skin surface is necessary to make the guiding graphics appear correctly registered to the patient.

In [Glo03] a system is described which utilizes infrared and visible lasers to project guiding graphics, such as entry points and surgical plans, directly onto the patient. Since it is possible to also measure the location of projected infrared laser spots, the system can assist in registration and ensure the accuracy of the display.

A similar approach, where guiding graphics is projected directly onto the patient, is described in [Hop02, Wör01, Hop03]. Here a video projector displays the surgical plan directly on the patient. Furthermore 3D information of the patient's surface can be gained by structured light techniques and two video cameras that evaluate the projected patterns on the patient.

Chapter 3

RAMP—A Novel System for AR Guidance

Developing and engineering an augmented reality system that is geared for interventional image guidance requires a design process that tackles the fundamental challenges of augmented reality as described in Chapter 2.2 in the realm of its potential application. Proper visualization, tracking, and registration techniques have to be developed and integrated into a hardware and software architecture that merges medical imaging information with the real view of patient and medical instruments. Main goals are to build a genuine AR display system, and optimize it for real-time operation, minimum latency, and high accuracy, and especially to prove its overall acceptance and value in the clinical field.

This chapter describes a research prototype AR system that we developed at Siemens Corporate Research [Vog06, Vog03, Sau00] and evaluated together with clinical collaborators to iteratively refine its accuracy, reliability, and functionality. Its technical and pre-clinical evaluation will be reported in Chapter 4 and Chapter 5. This prototype system has been presented as the RAMP system to the research community, where *RAMP* stands for “Reality Augmentation for Medical Procedures.”

3.1 System Design

Various approaches to augmented reality for interventional image guidance were presented in Chapter 2.7. Many attributes of a system are directly related to the type of the chosen display. Real-time performance, latency, and the representation of the augmented view differ for video vs. optical, head-mounted vs. stationary, and stereoscopic vs. monoscopic display technologies.

3.1.1 Video See-Through Approach

The RAMP system is built around a head-mounted video see-through display, that we assembled in our lab. A head-mounted display has the potential to provide a dynamic, stereoscopic viewpoint to the user of the system and as such presents a very intuitive access to the augmented scene. Even though the display as described in the following is bulky and relatively heavy, the underlying idea has been to optimize all other attributes of the system, such as accuracy, display resolution, latency, real-time performance, acceptance of the fused view, etc., to provide a base for experimental evaluation of the methodology of video see-through AR in the medical arena. If this technology proves itself successful and well accepted for interventional image guidance, minimization of the display would move to the foreground.

We chose a video see-through approach even though it is technically more demanding than a conventional optical see-through approach with semitransparent displays. An optical see-through system would not require that the user wears two cameras on his head, it would not require the processing of the corresponding pair of video streams, and it would still provide the "real part" of the augmented scene with unmatched resolution. There are several benefits of the video see-through approach that can justify the extra effort. They all stem from the fundamentally different way in which the augmented view is generated. With an optical see-through system the merging of real view and virtual objects takes place ultimately only in the user's eye. A video see-through system composites the augmented view already in the computer and allows much more control over the result. Chapter 2.4.3 compares the main properties of optical and video see-through approaches. The driving factors for us to chose a head-mounted video see-through approach are the following major benefits of these types of systems:

- ***Complete visualization control:*** There is control over the visualization of the real scene, which can, e.g., be dimmed down or made brighter to optimize the contrast.

- ***Synchronization:*** There is control over the timing of the real scene. An optical see-through system has an intrinsic time lag between the presentation of the real scene (which is immediate) and the presentation of the virtual objects from the corresponding viewpoint (for which first the head pose has to be determined before the virtual objects can be rendered and finally displayed). With a video see-through system real and virtual images can be synchronized (at the expense of a slight delay of the real scene), so that the virtual objects appear as true static parts of the scene.

- ***Precision:*** Calibration, as described in Chapter 2.5.3, is necessary to register the virtual objects correctly with respect to the real scene. The calibration of an optical see-through system with semitransparent displays requires subjective user interaction, as it is only the user who sees the composite image. The calibration of a video see-through system, however, is based on captured images and can be performed objectively in the computer. Calibration is much more precise here. In fact, we have used an early prototype of our video see-through system to quantitatively evaluate a method for optical see-through system calibration [Gen00]. Optical microscopes equipped for augmented reality visualization get around this problem by combining real and graphic views in an intermediate image plane (see Chapter 2.7.1). Here, the registration is fixed in an objective way, independent of the user.

- ***Reliability:*** The subjectivity of the optical see-through approach with semitransparent displays can be a safety problem. If the head-mounted display shifts relative to the user's eyes after the calibration procedure, the registration is no longer valid, but there is no objective warning signal. For example in a medical scenario where a surgeon is guided by the augmentation, such a misregistration would not be tolerable. A shift of a video see-through display on the user's head, however, does not influence the registration of the augmented view.

- ***Sharing:*** An augmented view of a video see-through system can be shared with many people. Hence, it also provides straight forward access to documentation, training, remote expert consultation, etc.

3.1.2 HMD Hardware Components

The centerpiece of the system is the stereoscopic video see-through head-mounted display shown in Figure 3.1. A Kaiser ProView XL35, which is a virtual reality HMD that is mostly used for military applications such as flight simulators, is equipped with two Panasonic GP-KS1000 miniature color cameras that serve as the user's artificial eyes. The Kaiser HMD has XGA (1024×768) resolution, the highest resolution among the commercially available LCD-based HMDs. In stereo mode, the HMD is driven with two independent XGA signals. The Panasonic cameras have one of the highest resolutions among single CCD-chip miniature color video cameras; with 910,000 imaging elements on the CCD chip, they contain about twice as many pixels as other comparable cameras in the small lipstick format. Although the incoming video is captured with VGA (640×480) resolution (and thus has to be scaled to XGA), the higher resolution of the display still provides great benefit for the quality of the augmenting graphics. The lenses of both cameras have a focal length of about 15.8mm, which results in a diagonal

Figure 3.1: *Top:* Stereoscopic video see-through head-mounted display with XGA resolution. *Bottom:* Camera triplet comprising a stereo pair of cameras to capture the scene and a dedicated tracker camera with IR LED illuminator.

field-of-view (FOV) of 30.18˚. Since the HMD provides a diagonal FOV of 35˚, we have a slight magnification of 1.16. We found that this magnification nicely compensates for the fact that the cameras are mounted farther away (i.e., on the head) from the workspace than the actual displays of the HMD are (i.e., in front of the eyes). Since in all our pre-clinical setups the user concentrates on a specific and locally confined task while using the AR system, the FOV of 35˚does not seem to be a limiting factor. The peripheral vision, around the displays of the HMD, is only relevant to give the user some degree of awareness of his environment.

A special feature of the system is a third head-mounted camera, a monochrome Sony XC-77RR, used for tracking. The Sony camera is equipped with a visual-cut/infrared-pass optical filter to make it sensitive only for wavelengths in the near-infrared. It works in conjunction with a set of IR LEDs that are placed around the lens and illuminate retroreflective markers in the scene. Our first illuminator provided a continuous light output. Later we replaced it with an IR flash designed for us by the company A.R.T. [A.R02]. The flash is synchronized with the tracker camera and allows us to select a fast speed of its electronic shutter. Early experiments in a neurosurgical operating room had shown us that the surgical halogen lamps above the surgical table emit a strong near-infrared light as well. Therefore the short exposure time of only 0.25 ms of the flash/camera combination efficiently suppresses background light in the images of the tracker camera, even when the scene is lit with strong incandescent or halogen lamps. Chapter 3.2 will provide the details on our tracking method. The camera triplet with IR flash is shown in the lower part of Figure 3.1.

3.1.3 AR Hardware Architecture

Processing the three incoming video streams with 30 Hz frame rate and realizing a correspondingly fast stereoscopic graphics output in XGA (1024×768) resolution puts high demands on the computational bandwidth and architecture of the AR system. We built our first AR system by using three networked SGI Visual Workstations [Sau00, Sau01b]. Figure 3.2 outlines the system structure for that setup on the left. One SGI PC processed the video input of the tracker camera to calculate the pose of the camera triplet. The other two SGI PCs used this pose information to augment the video streams of the scene cameras with 3D graphics (i.e., medical data) for the left and the right display of the HMD. The communication between the tracker PC and the two rendering PCs was done via an ethernet link with the standard TCP/IP protocol.

Recent advances in graphics hardware, memory bandwidth, and computing power of standard PCs made it possible to transfer the technology of our AR system to a single PC with dual monitor graphics support. The right side of Figure 3.2 shows the structure of the single PC system. The graphics system (NVIDIA GeforceFX 5950) is capable of generating a single monoscopic AR image in 5–15 milliseconds, depending on the complexity of graphics to be rendered on top of the video image, which itself is uploaded into the graphics memory and scaled from VGA

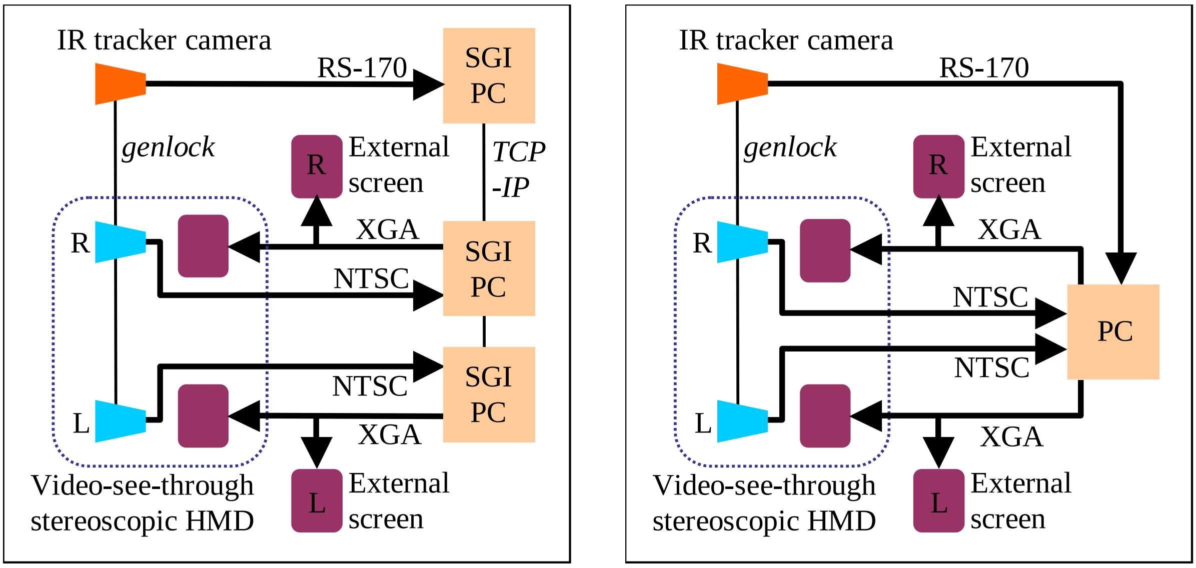


Figure 3.2: Hardware structure of the AR system. *Left:* The system which we first introduced in [Sau00] is based on 3 dedicated SGI visual workstations. *Right:* The new single PC based AR system with 3 NTSC camera inputs and 2 XGA outputs outperforms the 3 SGI workstation solution.

(640×480) to XGA (1024×768) format as part of the rendering process. This speed is necessary to achieve real-time performance with 30 frames/second, where each stereo image has to be processed and visualized within 33 milliseconds.

An important feature of the video see-through system is that the three cameras are genlocked so that the video streams arrive at the computer in a synchronized way. We preserve the synchronization by time-stamping the incoming frames and by using the pose information deduced from a particular tracker camera frame to augment the scene camera frames that have been captured at the same time. In that way, there is no time-lag between video and graphics. The graphics appears firmly anchored in the augmented video images, which contributes to a very believable perception of the augmentation.

3.1.4 AR Software Architecture

Augmented reality is a multi-disciplinary field, covering the areas of computer vision, computer graphics, human-computer interfaces, real-time systems, and more. Correspondingly, it requires a powerful, highly flexible, but also extremely efficient underlying software architecture. In a flexible way we would like to be able to change AR hardware setups—like types and numbers of video sources, tracking devices, graphics hardware, displays, and input devices—or employ

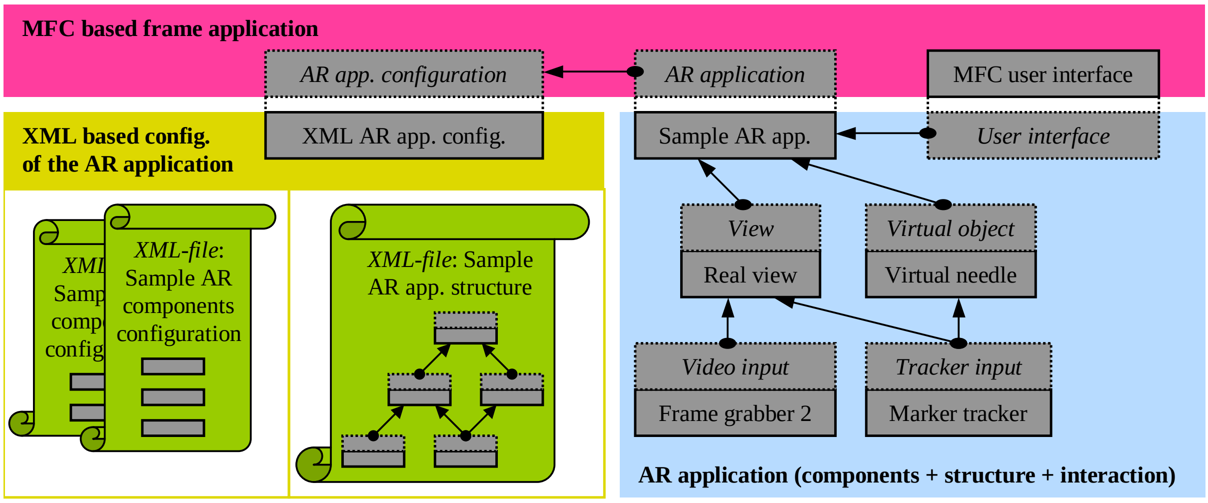


Figure 3.3: Software component structure of the extensible AR system. The components provide true information encapsulation by communicating through signals and slots. The structure of AR applications is defined in XML files as well as the attributes of the incorporated components. The AR demands of a highly flexible, but also extremely efficient software architecture were taken as a starting point for the development of this C++ library.

various user interaction strategies, and also maintain different AR research environments and prototype installations at the same time. These requirements can only be fulfilled with a good abstraction hierarchy during the development of reusable AR software components. The basic concepts are topics of ongoing research [Bau01, Sch98, Uch02, Bil02].

We designed a modular, extensible, and very efficient component based AR software library. The major issues were:

- Easy integration of new components into the system,
- Ability to configure each component independently,
- Defining AR application structures on a higher level.

With *component* we generally mean some piece of code that fulfills a specific task, e.g. providing an interface to a certain hardware device like a camera, a video file on the hard drive, or an external tracker, executing a computer vision algorithm, rendering some part of a virtual 3D scene, or dealing with user interactivity. The definition of an *AR application structure* should be based on a certain abstraction level of the incorporated components.

The principal idea behind this programming environment is depicted in Figure 3.3. The right side of the chart illustrates how the components of an AR application interact with each other by using *signals* and *slots*, which permits real information encapsulation, a fundamental concept of component programming. For example, if a component changes states it can emit according signals without knowing which other components are receiving them and thus act on them. In [Pfi96] it has been shown why object-oriented programming itself is insufficient to ensure the design of true components. Reference [Szy02] gives further details about the techniques of component-oriented programming. In Figure 3.3 a signal/slot connection is represented by an arrow from the component emitting the signal to the component receiving the signal (with its slot). The abstract interface to the signals of a component is indicated by a broken framed box, e.g., ‘VideoInput’. The actual implementation of the component is specific and symbolized by a closed framed box, e.g., ‘Frame grabber 2’. On the one hand, this level of abstraction allows an independent generation and configuration of each specific component, and on the other hand allows the definition of the AR application structure on an abstract level.

Another key element is the way of configuring the component attributes and the application structure by using XML. The actual components and the frame application are efficiently implemented in C++. On the one hand, XML serves as a meta-language to easily create and maintain AR applications and on the other hand is used to configure each single component. A big advantage of XML is that there are many external tools available to edit those files.

3.1.5 Stereoscopic AR Scene Composition

We want to superimpose graphics onto the video images so that the graphics objects appear to be a natural part of the scene. Merging the left and right video views of the real world with virtual objects, requires to model both real cameras as two virtual cameras inside the virtual 3D scene. We use the intrinsic and extrinsic parameters of the scene cameras (Chapter 2.5.2, Chapter 2.6) to render a stereoscopic image of the virtual world, which is blended onto the incoming stereo images of the scene cameras. We found that we can neglect the radial distortion of our scene cameras and simply model the two virtual cameras as pinhole cameras.

We implement the rendering of the augmented scenes with OpenGL, what we display is the view of the OpenGL camera. The video is incorporated as a texture map onto a “virtual screen” in the background. This virtual screen is fixed within the coordinate system of the virtual camera. For correct registration we have to put the virtual camera in the right place, based on the tracked pose of the real camera. We also implemented a digital zoom for the augmented scene by simply varying the field of view of the viewing frustum of the OpenGL camera. Although an optical zoom would provide a better quality of the zoomed video view, the digital zoom still allows the user to gain more details from the augmenting graphics.

3.2 Single Camera Marker Tracking

Tracking is an enabling technology for augmented reality applications. Commercial tracking systems are generally based either on optical or magnetic technologies (Chapter 2.6). Optical systems utilize two or multiple cameras to estimate 3D marker positions through triangulation. The use of a single camera for tracking is known to the computer vision community but in the context of an dedicated marker tracker in the near infrared spectrum for the pose estimation of head-mounted displays and surgical instruments it has been unexplored.

3.2.1 Head Mounted Tracker Camera

The decision to develop our own tracking system with a single head-mounted camera was based on several considerations. Mounting the tracker camera on the user's head (also referred to as an *inside-out* system [Wan90]) helps with the line-of-sight restriction of optical tracking; the user cannot step into the field of view (FOV) of the tracker camera (though the user still can occlude markers with his hands). Placing a forward looking tracking camera on the head is also optimal for the perceived accuracy of the augmentation, as the sensitivity of the tracker camera to registration errors is matched to the user's sensitivity to perceive these errors [Hof98, Har93]. Movements along the optical axis are tracked with a lower accuracy than transverse movements, but at the same time a depth error in the position of a virtual object is also less perceptible than a lateral error. In other words: when scene and tracker camera look in the same direction, the camera detects just what the user can see.

Why introduce an extra tracker camera at all? Of course, one could just use the two scene cameras for tracking. That would also promise the best augmentation accuracy. But it would require equipping the workspace volume with markers or features that can be robustly detected. Since the field of view of the scene cameras is very limited, the user can easily occlude these markers or features, and the markers can easily get in the user's way. Our specialized infrared tracker camera with fish-eye lens instead provides a solution, which puts no constraints on the content of the workspace volume and which simplifies the tracking task for real-time operation.

3.2.2 Marker Set Design for Head Tracking

We distinguish two sets of markers that are tracked by the head-mounted tracker camera. The first set is used for head tracking where the markers are arranged on a stationary frame around the workspace. It is shown in Figure 3.4. Marker-based tracking usually utilizes either spherical

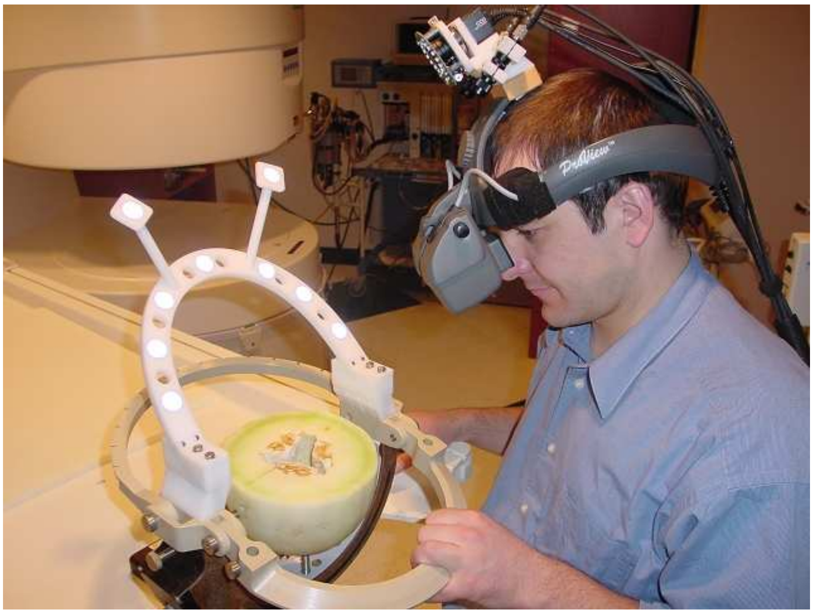


Figure 3.4: Marker configuration for head tracking. A cantaloupe is placed into the head-clamp of a neurosurgical iMR operating room and viewed with the video-see-through HMD. Head tracking works in conjunction with nine retro-reflective markers which are reproducibly attached to the top of the head-clamp.

or disc-shaped markers. We chose disc-shaped markers since they are easy to produce by punching them out of a sheet of retro-reflective material. The 3D coordinates of the markers, relative to each other, are measured with the commercially available multi-camera tracking system *ARTtrack1* [A.R02].

This marker configuration makes good use of the FOV of the tracker camera since head tracking is only required when the user actually looks at the workspace. For this reason, these markers can extend over a sizable part of the image, yielding good tracking accuracy [Har93]. It is also noteworthy to mention that this frame marker set is designed to extend around the workspace as much as possible, providing good accuracy in the position of the target points all over the workspace.

3.2.3 Marker Cluster Design for Instrument Tracking

The second set is a relatively small cluster of markers, which can be attached to hand-held tools or instruments. The right part of Figure 3.5 shows a biopsy needle with an attached marker cluster. For practical reasons, this marker cluster should be small. But for a single camera tracking system, miniaturizing the marker set could considerably decrease tracking accuracy [Har00].

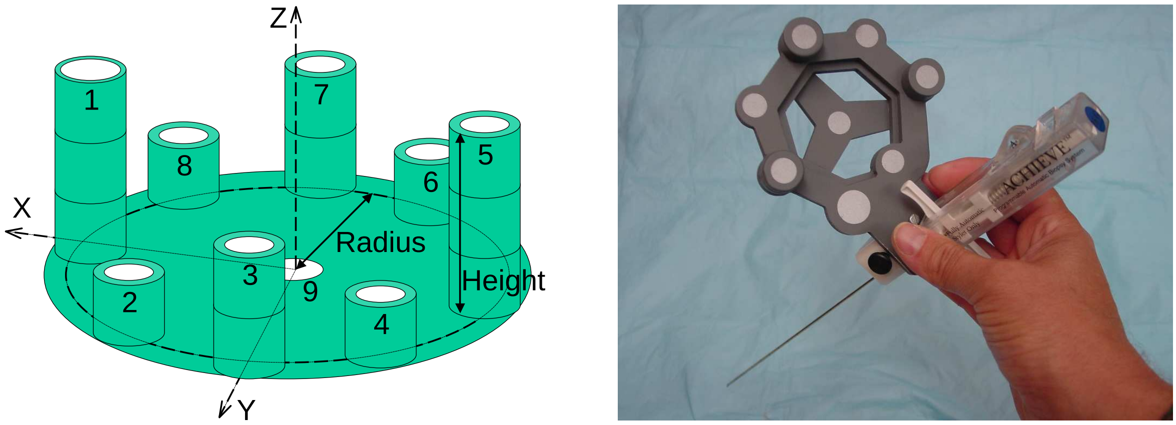


Figure 3.5: Marker configuration for instrument tracking. *Left:* Schematic representation of the design of the introduced family of marker clusters. The markers (in white) are rigidly mounted on little posts that are arranged in a circular shape. The circle base lies in the x-y plane of the marker cluster coordinate system. *Right:* A biopsy needle which is equipped with a set of retro-reflective markers. This marker cluster is tracked by the head-mounted tracker camera to estimate the position and orientation of the needle. This permits the AR system to augment the scene with guiding graphics, such as a representation of the needle trajectory inside the patient.

This limitation is less crucial in a multi-camera tracking system. Furthermore, the distribution of markers in a cluster is constrained by the fact that all of them must be visible in the camera image over a wide range of cluster poses without possible occlusions.

We consider marker clusters with markers located on the perimeter of a circle, in various heights (z-coordinate) relative to the (x-y-) plane of the circle. Additionally, one marker is placed in the center of the circle. On the one hand, a symmetric marker distribution presumably leads to a symmetric error distribution in the x- and y-direction of the marker cluster coordinate system. On the other hand, this cluster design allows a simple identification of the markers in the camera image, since a view from the top at this marker cluster reveals its circular structure. For labeling we have one marker with a greater diameter, whereas all the other markers have the same size. A schematic representation of this marker configuration is illustrated in the left of Figure 3.5. We choose three parameters to characterize the introduced family of marker clusters:

- the number of markers,
- the radius of the circle, and
- the maximum height of a marker compared to the plane of the circle.

The last item represents the degree of non-coplanarity and indicates how the markers are spatially distributed in the z-direction.

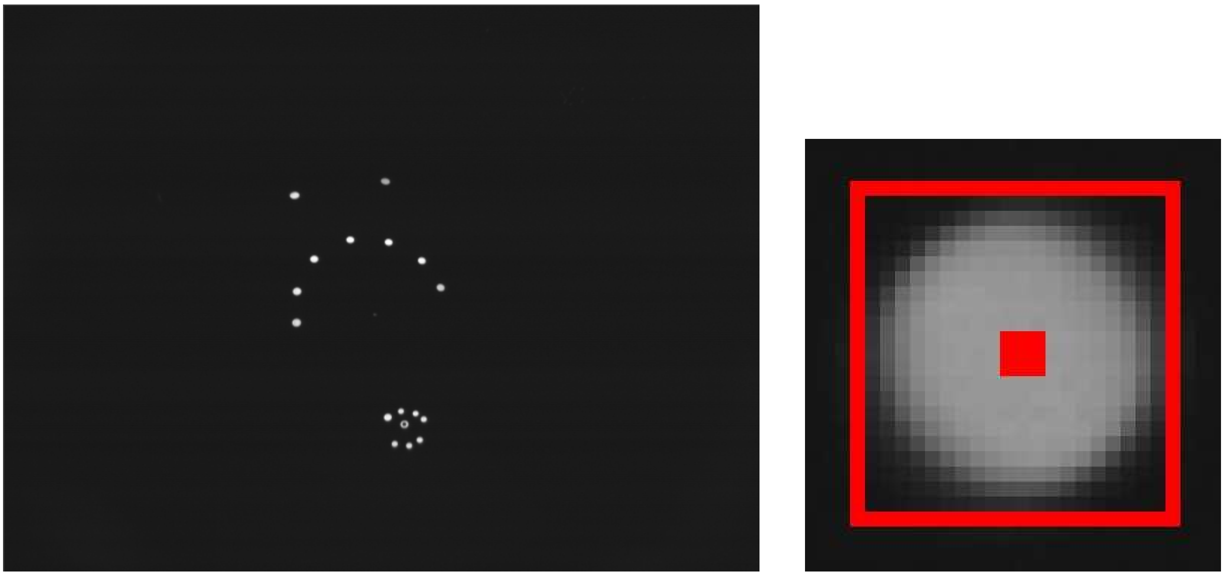


Figure 3.6: *Left:* Image of the tracker camera while tracking the markers that frame the workspace and a marker cluster that is attached to a hand-held tool. The wide-angle lens ($\approx 110^\circ$) ensures a large tracking range, but also puts high demands on the sub-pixel precision of the estimated marker centers. *Right:* Zoomed view of an imaged marker. The marker center is determined by calculating the moments on the gray values in a bounding box.

In Chapter 4.2 and [Vog02] we optimize this marker cluster to achieve minimal variation in the estimated pose parameters, i.e., to minimize any visible jitter in the final augmented video view. We found that a cluster with 8 markers, 45mm radius, and 30mm cluster height provides stable tracking results.

3.2.4 Marker Detection Strategy

Figure 3.6 shows a typical image of the tracker camera on the left. The eight smaller markers in the lower part of the image belong to the tracked marker cluster (Chapter 3.2.3) that is attached to a biopsy needle, whereas the nine larger markers are the ones that frame the workspace (Chapter 3.2.2). The retro-reflective markers, which are illuminated by near-infrared light, give a very good contrast in the camera image and therefore allow an efficient and accurate localization. The strong infrared flash, which is synchronized with the shutter of the tracker camera, effectively suppresses distracting background light from other infrared emitting light sources in the surrounding environment. Additionally, due to the short exposure time of only 0.25 milliseconds, any motion blurring of the retro-reflective markers during fast movements of the HMD is greatly prevented.

We use disc-shaped markers, thus the projected shapes in the camera image are elliptical. At first we segment the markers in the binarized image with a connected component analysis by tracing the image for bright pixels [Bal82, pg. 151]. Figure 3.6 depicts on the right side a magnified view of a marker in the tracker camera image and its circumscribing bounding box. We extract the 2D coordinates of the projections of the marker centers by computing the centroid of each projected marker with sub-pixel precision from moments of the original gray-value image. The moment of order $(p+q)$ for a function $f(x,y)$ of two variables is (e.g., [Nie90, pg. 100])

$$m_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^q f(x,y)\,dx\,dy \tag{3.1}$$

and the center of gravity is defined by

$$x_\mathrm{c} = \frac{m_{10}}{m_{00}}\,, \quad y_\mathrm{c} = \frac{m_{01}}{m_{00}}\,. \tag{3.2}$$

Assuming that the retroreflective material ensures a homogeneously distributed brightness along the extent of the marker, the calculation of the center of gravity of the gray values in the bounding box in the camera image will result in a sub-pixel precise estimation of the projected center of the real marker. To compensate for background light we subtract a constant $b \in \mathbb{R}$ from each pixel before the calculation of the center of gravity. The background brightness can be assumed locally constant for each marker. We simply estimate b by averaging all pixels on the boundary of the bounding box of the marker. If $w, h \in \mathbb{N}$ are the width and height of the bounding box and $x_0, y_0 \in \mathbb{N}$ are the coordinates of its left upper corner in the image and, furthermore, $g_{x,y} \in 0, 1, \ldots, 255$ describe the gray values in the digitized camera image at coordinates (x,y), then the center of gravity for each marker in the image can be calculated as:

$$x_\mathrm{c} = \frac{\sum\limits_{x=x_0}^{x_0+w-1}\left(\sum\limits_{y=y_0}^{y_0+h-1}(g_{x,y}-b)\right)x}{\sum\limits_{x=x_0}^{x_0+w-1}\sum\limits_{y=y_0}^{y_0+h-1}(g_{x,y}-b)} = \frac{\sum\limits_{x=x_0}^{x_0+w-1}\left(\sum\limits_{y=y_0}^{y_0+h-1} g_{x,y}\right)x - \frac{1}{2}(w-1)whb}{\sum\limits_{x=x_0}^{x_0+w-1}\sum\limits_{y=y_0}^{y_0+h-1} g_{x,y} - whb}\,, \tag{3.3}$$

$$y_\mathrm{c} = \frac{\sum\limits_{y=y_0}^{y_0+h-1}\left(\sum\limits_{x=x_0}^{x_0+w-1}(g_{x,y}-b)\right)y}{\sum\limits_{x=x_0}^{x_0+w-1}\sum\limits_{y=y_0}^{y_0+h-1}(g_{x,y}-b)} = \frac{\sum\limits_{y=y_0}^{y_0+h-1}\left(\sum\limits_{x=x_0}^{x_0+w-1} g_{x,y}\right)y - \frac{1}{2}(w-1)whb}{\sum\limits_{x=x_0}^{x_0+w-1}\sum\limits_{y=y_0}^{y_0+h-1} g_{x,y} - whb}\,. \tag{3.4}$$

The right sides of (3.4) and (3.4) are optimized for real-time performance since they don't require

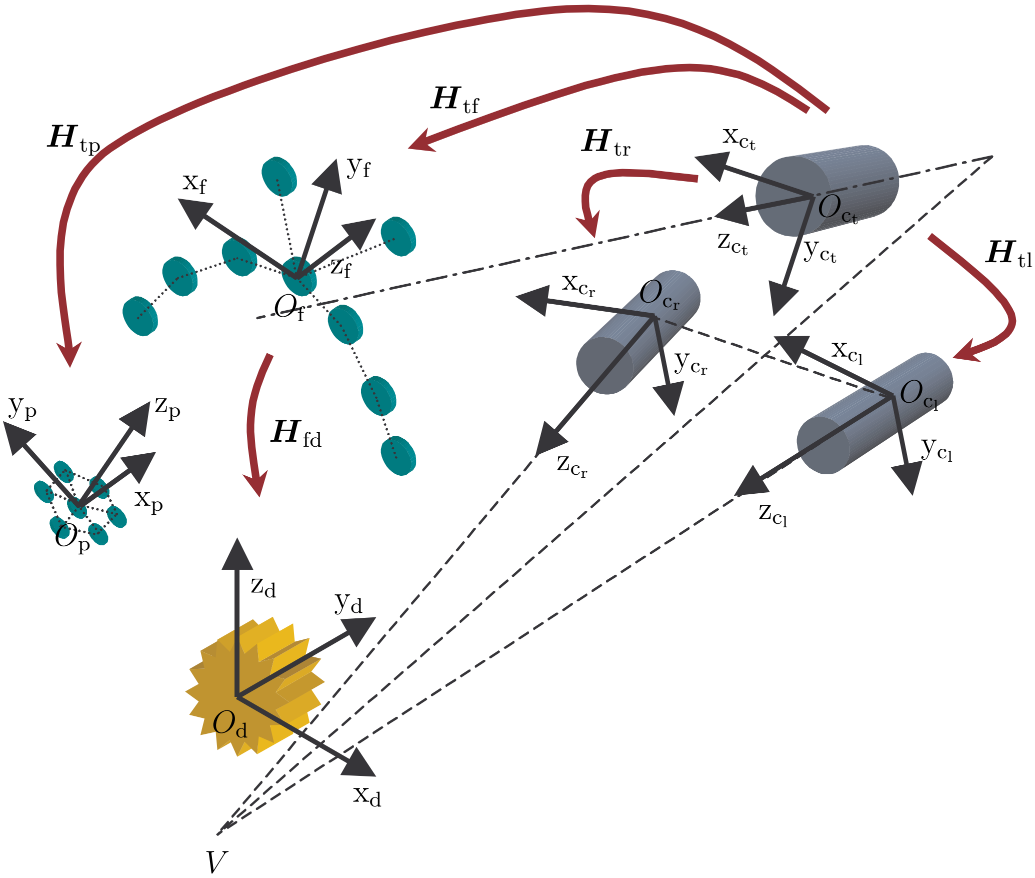


Figure 3.7: Coordinate transformations which are estimated by the system. The rigid body transformations $\boldsymbol{H} \in \mathbb{R}^{4\times4}$ map points given in projective coordinates $\tilde{\boldsymbol{p}} \in \mathcal{P}^3$ between the different 3D coordinate systems. The three cylinders symbolize the three cameras, the small discs the markers, and the gear the in-situ visualized medical patient data.

to actually subtract the background brightness from each pixel in the bounding box. Two simple subtractions (in the nominator and denominator of the fraction) lead to the same results.

It should be noted that no further geometrical assumptions, such as for ellipse-fitting methods or similar methods in camera calibration, are used since the retroreflective properties of the circular markers ensure a homogeneous reflection of the infrared light toward the camera and as such the center of gravity of the incoming light will coincide with the projected center of the circle. The complete marker detection method results in center coordinates of n markers in the tracker camera image, which are written in homogeneous coordinates $\{\tilde{\boldsymbol{q}}_{\mathrm{i_t},\mu} \in \mathcal{P}^2 \mid \mu = 1, 2, \ldots, n\}$. These results are used for the purpose of calibration and tracking as explained later.

3.2.5 Camera and Scene Geometry

Figure 3.7 shows a sketch of all the coordinate systems and transformations involved in the AR setup. All three cameras have a camera centered right handed coordinate system assigned where the $\mathrm{z_c}$-axis is aligned with the optical axis. The two (color) scene cameras (left: $\mathrm{c_l}$; and right: $\mathrm{c_r}$) have a baseline $\overline{O_{\mathrm{c_l}}O_{\mathrm{c_r}}}$ of about 6.5 cm, corresponding to an average interpupillary distance. We focused the lenses to a distance of about 60 cm, which is comfortable working distance. For an optimal overlap of the stereo images, we correspondingly fixed both cameras so that their optical axes intersect at this same distance. Since these two cameras work as artificial eyes for the user, it is important to carefully adjust the locations and rotations (around the optical axes) of both cameras relative to each other before fixating them. In particular, the $\mathrm{x_{c_l}}$-$\mathrm{z_{c_l}}$-plane of the left camera and the $\mathrm{x_{c_r}}$-$\mathrm{z_{c_r}}$-plane of the right camera must geometrically be the same and $\triangle(O_{\mathrm{c_r}}, V, O_{\mathrm{c_l}})$ must be an isosceles triangle (with basis $\overline{O_{\mathrm{c_l}}O_{\mathrm{c_r}}}$)—a configuration that simulates both human eyes when focusing at a point V.

The tracker camera ($\mathrm{c_t}$) is mounted above the midpoint of the baseline of the scene cameras in a way that the $\mathrm{y_{c_t}}$-$\mathrm{z_{c_t}}$-plane contains the point V and is orthogonal to the plane of the stereo cameras (the $\mathrm{x_{c_l}}$-$\mathrm{z_{c_l}}$-plane of the left camera and the $\mathrm{x_{c_r}}$-$\mathrm{z_{c_r}}$-plane of the right camera). It furthermore looks upward from this plane under an angle of 20°, which is adapted to the geometry of the augmented workspace with its framing markers beside and above the area of interest (e.g., Figure 3.4).

The rigid body transformations (given by the extrinsic parameters) between the tracker and the scene cameras are denoted by the matrices $\boldsymbol{H}_{\mathrm{tl}}, \boldsymbol{H}_{\mathrm{tr}} \in \mathbb{R}^{4\times4}$. If the homogeneous coordinates of a point P in the tracker camera coordinate system are given by $\tilde{\boldsymbol{p}}_{\mathrm{c_t}} \in \mathcal{P}^3$, then the homogeneous coordinates of that point P in the left camera coordinate system $\tilde{\boldsymbol{p}}_{\mathrm{c_l}} \in \mathcal{P}^3$ and in the right camera coordinate system $\tilde{\boldsymbol{p}}_{\mathrm{c_r}} \in \mathcal{P}^3$, are given by

$$\tilde{\boldsymbol{p}}_{\mathrm{c_l}} = \boldsymbol{H}_{\mathrm{tl}}\, \tilde{\boldsymbol{p}}_{\mathrm{c_t}} \quad \text{and} \quad \tilde{\boldsymbol{p}}_{\mathrm{c_r}} = \boldsymbol{H}_{\mathrm{tr}}\, \tilde{\boldsymbol{p}}_{\mathrm{c_t}} \,. \tag{3.5}$$

The workspace, where real objects will be augmented with virtual ones, is framed by a set of markers, whose 3D coordinates are known in a frame coordinate system $(O_{\mathrm{f}}, \mathrm{x_f}, \mathrm{y_f}, \mathrm{z_f})$. The rigid body transformation $\boldsymbol{H}_{\mathrm{tf}} \in \mathbb{R}^{4\times4}$ transforms the homogeneous coordinates $\tilde{\boldsymbol{p}}_{\mathrm{c_t}} \in \mathcal{P}^3$ of a point P in the tracker camera coordinate system to the homogeneous coordinates $\tilde{\boldsymbol{p}}_{\mathrm{f}} \in \mathcal{P}^3$ in the frame coordinate system:

$$\tilde{\boldsymbol{p}}_{\mathrm{f}} = \boldsymbol{H}_{\mathrm{tf}}\, \tilde{\boldsymbol{p}}_{\mathrm{c_t}} \,. \tag{3.6}$$

The object in the real world which will be augmented with medical data, also has an own coordinate system attached $(O_\mathrm{d}, \mathrm{x_d}, \mathrm{y_d}, \mathrm{z_d})$ and $\boldsymbol{H}_\mathrm{fd} \in \mathbb{R}^{4\times4}$ represents the rigid body transformation of the homogeneous coordinates $\tilde{\boldsymbol{p}}_\mathrm{f} \in \mathcal{P}^3$ of a point P in the frame coordinate system to the homogeneous coordinates $\tilde{\boldsymbol{p}}_\mathrm{d} \in \mathcal{P}^3$ in the medical data coordinate system:

$$\tilde{\boldsymbol{p}}_\mathrm{d} = \boldsymbol{H}_\mathrm{fd}\, \tilde{\boldsymbol{p}}_\mathrm{f} \,. \tag{3.7}$$

To augment a certain point P in the real world with computer graphics, whose coordinates are given in the medical data coordinate system, one has to apply the following transformations to calculate the coordinates of this point in the left and right camera coordinate systems:

$$\tilde{\boldsymbol{p}}_{\mathrm{c_l}} = \boldsymbol{H}_\mathrm{tl}\, \boldsymbol{H}_\mathrm{tf}^{-1}\, \boldsymbol{H}_\mathrm{fd}^{-1}\, \tilde{\boldsymbol{p}}_\mathrm{d} \,, \tag{3.8}$$

$$\tilde{\boldsymbol{p}}_{\mathrm{c_r}} = \boldsymbol{H}_\mathrm{tr}\, \boldsymbol{H}_\mathrm{tf}^{-1}\, \boldsymbol{H}_\mathrm{fd}^{-1}\, \tilde{\boldsymbol{p}}_\mathrm{d} \,. \tag{3.9}$$

Since both scene cameras are modeled as virtual cameras, with the same intrinsic camera parameters as the real ones (following Chapter 2.5.2), these camera coordinates are used to place a virtual object into the scene at the location of point P during the operation of the AR system.

To be able to track a hand-held tool or pointer, we attach a cluster of retro-reflective markers to the tool or pointer, whose 3D coordinates are known in a pointer coordinate system $(O_\mathrm{p}, \mathrm{x_p}, \mathrm{y_p}, \mathrm{z_p})$. To be able to visualize the tool or pointer inside the medical data set, we have to transform its known coordinates from the pointer coordinate system to the medical data coordinate system. Equation

$$\tilde{\boldsymbol{p}}_\mathrm{d} = \boldsymbol{H}_\mathrm{fd}\, \boldsymbol{H}_\mathrm{tf}\, \boldsymbol{H}_\mathrm{tp}^{-1}\, \tilde{\boldsymbol{p}}_\mathrm{p} \tag{3.10}$$

transforms the homogeneous coordinates $\tilde{\boldsymbol{p}}_\mathrm{p} \in \mathcal{P}^3$ of a point P from the pointer coordinate system to the homogeneous coordinates $\tilde{\boldsymbol{p}}_\mathrm{d} \in \mathcal{P}^3$ in the medical data coordinate system, where $\boldsymbol{H}_\mathrm{tp} \in \mathbb{R}^{4\times4}$ represents the rigid body transformation from the tracker camera coordinate system to the pointer coordinate system.

Proper registration methods have to deliver all involved rigid-body transformations to utilize (3.8), (3.9), and (3.10). Chapter 3.2.6 describes how $\boldsymbol{H}_\mathrm{tl}$, $\boldsymbol{H}_\mathrm{tr}$, and the intrinsic parameters of all three cameras are estimated in the offline calibration phase. Chapter 3.2.7 gives details about the real-time tracking phase of the AR system, where $\boldsymbol{H}_\mathrm{tf}$ and $\boldsymbol{H}_\mathrm{tp}$ are estimated for each image of the tracker camera. The transformation $\boldsymbol{H}_\mathrm{fd}$ is specific to each interventional setup and its estimation will be described in the investigated studies.

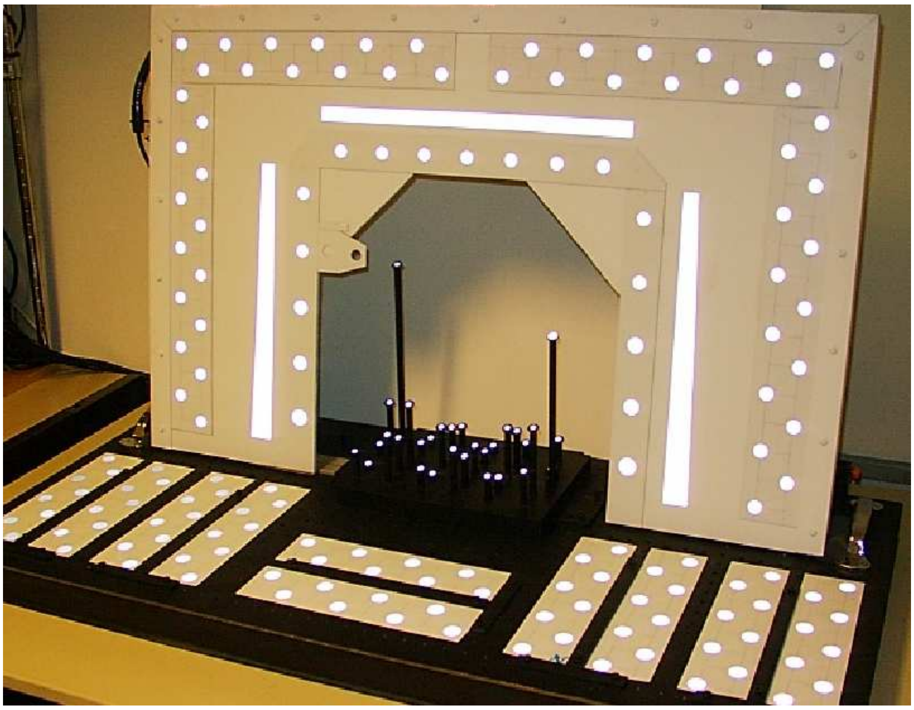


Figure 3.8: Calibration setup for calibrating the intrinsic and relative extrinsic parameters of tracking camera and scene cameras.

3.2.6 Offline Camera Calibration

We mount our three cameras—one tracking and two scene cameras—in a rigid configuration, and also fix their focal lengths. In an offline camera calibration step, we determine the intrinsic and relative extrinsic ($\boldsymbol{H}_{\mathrm{tl}}$, $\boldsymbol{H}_{\mathrm{tr}}$) camera parameters for this camera triplet.

One advantage of having a tracker camera look into the same direction as the scene cameras is that all cameras can see one single calibration object at the same time. This simplifies the calibration process and it makes it easier to achieve good accuracy.

Figure 3.8 shows our table-top 3D calibration object. It contains 179 disc-shaped markers. We measured the 3D coordinates of these markers with an optical tracking system from A.R.T. [A.R02]. The scene cameras have a much smaller field of view (diagonal $\approx 30^{\circ}$) than our wide-angle tracking camera (diagonal $\approx 110^{\circ}$). Correspondingly, the calibration object contains a set of small markers in the center (30) for the scene cameras and a set of larger markers in the periphery (149) for the tracking camera. Figure 3.9 illustrates the different FOVs for scene and tracker cameras.

If the homogeneous coordinates of a calibration point P are given in the coordinate system of the calibration object by $\tilde{\boldsymbol{p}}_{\mathrm{o}} \in \mathcal{P}^3$, then the homogeneous coordinates of the projected calibration

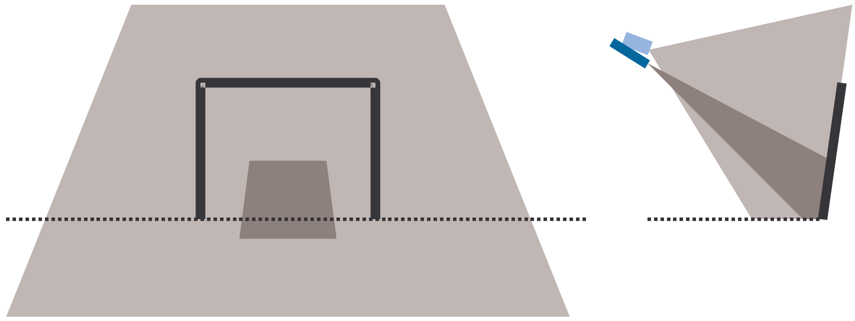


Figure 3.9: Simulated fields of view of tracker and scene camera (light and dark gray, respectively), as seen from the front and from the side. The broken line indicates the table level, the black frame indicates the location of the inner marker frame of the calibration pattern.

point in the images of the three cameras are given by:

$$\tilde{\boldsymbol{p}}_{\mathrm{u_l}} \cong \boldsymbol{A}_{\mathrm{l}}\, \boldsymbol{H}_{\mathrm{ol}}\, \tilde{\boldsymbol{p}}_{\mathrm{o}}\,, \quad \tilde{\boldsymbol{p}}_{\mathrm{i_l}} = \boldsymbol{d}(\tilde{\boldsymbol{p}}_{\mathrm{u_l}}, \{\kappa_{\mathrm{l}}\})\,, \tag{3.11}$$

$$\tilde{\boldsymbol{p}}_{\mathrm{u_r}} \cong \boldsymbol{A}_{\mathrm{r}}\, \boldsymbol{H}_{\mathrm{or}}\, \tilde{\boldsymbol{p}}_{\mathrm{o}}, \quad \tilde{\boldsymbol{p}}_{\mathrm{i_r}} = \boldsymbol{d}(\tilde{\boldsymbol{p}}_{\mathrm{u_r}}, \{\kappa_{\mathrm{r}}\})\,, \tag{3.12}$$

$$\tilde{\boldsymbol{p}}_{\mathrm{u_t}} \cong \boldsymbol{A}_{\mathrm{t}}\, \boldsymbol{H}_{\mathrm{ot}}\, \tilde{\boldsymbol{p}}_{\mathrm{o}}, \quad \tilde{\boldsymbol{p}}_{\mathrm{i_t}} = \boldsymbol{d}(\tilde{\boldsymbol{p}}_{\mathrm{u_t}}, \{\kappa_{\mathrm{t1}}, \kappa_{\mathrm{t2}}, \kappa_{\mathrm{t3}}\})\,, \tag{3.13}$$

where $\tilde{\boldsymbol{p}}_{\mathrm{u_l}}, \tilde{\boldsymbol{p}}_{\mathrm{u_r}}, \tilde{\boldsymbol{p}}_{\mathrm{u_r}} \in \mathcal{P}^2$ are the homogeneous coordinates in the undistorted images of the left, right, and tracker camera, respectively, and '$\cong$' indicates an equality up to a scale factor. $\boldsymbol{H}_{\mathrm{ol}}, \boldsymbol{H}_{\mathrm{or}}, \boldsymbol{H}_{\mathrm{ot}} \in \mathbb{R}^{4\times4}$ denote the rigid body transformations from the calibration object coordinate system to the corresponding camera coordinate system. The intrinsic camera parameters are represented by the matrices $\boldsymbol{A}_{\mathrm{l}}, \boldsymbol{A}_{\mathrm{r}}, \boldsymbol{A}_{\mathrm{t}} \in \mathbb{R}^{3\times4}$, respectively. We model radial lens distortion $\boldsymbol{d}$ and obtain the homogeneous coordinates in the distorted camera images as $\tilde{\boldsymbol{p}}_{\mathrm{i_l}}, \tilde{\boldsymbol{p}}_{\mathrm{i_r}}, \tilde{\boldsymbol{p}}_{\mathrm{i_t}} \in \mathcal{P}^2$ for the left, right, and tracker camera, respectively. For the scene cameras, we consider radial distortions up to the second order (parameters κ_{l} for the left and κ_{r} for the right camera). With their narrow fields of view, they show very little radial distortion. Since the tracker camera is equipped with a wide-angle lens, it exhibits strong radial distortion, which we model up to the sixth order, i.e., with three distortion parameters κ_{t1}, κ_{t2}, and κ_{t3}.

To calibrate the camera-triplet, we extended Tsai's calibration technique [Tsa87], which estimates the intrinsic, extrinsic and distortion parameters by using several 2D–3D point correspondences, in the following way.

At first the distortion and the intrinsic camera parameters are estimated for each camera separately by capturing several images (5–10) of the calibration object from different view points and applying Tsai's full algorithm to each of the images. This method utilizes 2D–3D correspondences from one image to calculate start values for the Levenberg-Marquart non-linear optimization routine, minimizing the reprojection error of all 3D points into that 2D image. This leads to a set of intrinsic, distortion, and extrinsic parameters for each calibration image of the camera. Since for a particular camera only the extrinsic parameters should change for each image, we subsequently utilize the Levenberg-Marquart optimization again, now minimizing the total reprojection error by including all 2D–3D correspondences of all calibration images of that camera, to optimize for exactly one set of intrinsic and distortion parameters (as well as one set of extrinsic parameters for each image). We use the previous results from Tsai's calibration method as start values for this optimization.

After repeating this extended calibration procedure for each of the three cameras, we get very robust estimates for $\boldsymbol{A}_{\mathrm{l}}, \boldsymbol{A}_{\mathrm{r}}, \boldsymbol{A}_{\mathrm{t}}$ and κ_{l}, κ_{r}, κ_{t1}, κ_{t2}, κ_{t3}. The camera matrices $\boldsymbol{A}_{\mathrm{l}}$ and $\boldsymbol{A}_{\mathrm{r}}$ will be used to model the virtual cameras during the real-time operation phase of the system. The reason for the increased robustness can be explained as follows. Different camera viewpoints toward the calibration object lead to different distributions of the calibration points in 3D space as seen from a camera-centered coordinate system. Even though our calibration objects exhibit calibration points that are distributed over all 3 dimensions, any bias towards a non-uniform distribution of those calibration points in 3D space and other possible systematic errors stemming from the construction of the calibration object will be eradicated by the multiple-view calibration approach of the intrinsic and distortion parameters of each camera. Furthermore, covering the range of view points of each camera during calibration that will also be covered during the real-time use of the AR system with regard to the tracked marker clusters will result in more precise overlays in general, as the calibration does not favor a particular viewing direction towards the disc-shaped markers as well as their 3D distribution.

To estimate the relative extrinsic parameters of the camera-triplet, one image is captured with each camera, while the camera-triplet is in a fixed frontal position to the calibration object. These three images are used to estimate only the extrinsic camera parameters ${}^{\hat{o}}\boldsymbol{H}_{\mathrm{ol}}$, ${}^{\hat{o}}\boldsymbol{H}_{\mathrm{or}}$, ${}^{\hat{o}}\boldsymbol{H}_{\mathrm{ot}}$ for this particular view $\hat{o}$ by using the formerly determined distortion and intrinsic camera parameters. The constant relative extrinsic camera parameters, which we need for (3.8) and (3.9), are now given by

$$\boldsymbol{H}_{\mathrm{tl}} = {}^{\hat{o}}\boldsymbol{H}_{\mathrm{ol}} {}^{\hat{o}}\boldsymbol{H}_{\mathrm{ot}}^{-1} \quad \text{and} \quad \boldsymbol{H}_{\mathrm{tr}} = {}^{\hat{o}}\boldsymbol{H}_{\mathrm{or}} {}^{\hat{o}}\boldsymbol{H}_{\mathrm{ot}}^{-1} . \tag{3.14}$$

The reason why we perform a joint estimation of all camera parameters for each camera separately but not for the triplet altogether (i.e., including $\boldsymbol{H}_{\mathrm{tl}}$ and $\boldsymbol{H}_{\mathrm{tr}}$) is because the images that are taken of the calibration object are much better when the camera viewpoint can be optimized for each camera independently. Thus, the intrinsic camera parameters, which are independent for each camera anyway, can be estimated more precisely when the image coverage of calibration points and the variety of viewpoints is optimized for each camera separately. Taking simultaneously images of the calibration object with all three cameras requires a certain viewpoint constellation, which delivers images that are good enough to estimate $\boldsymbol{H}_{\mathrm{tl}}$ and $\boldsymbol{H}_{\mathrm{tr}}$, but which would rather influence the intrinsic camera parameter estimation in a negative way if they were used for an overall joint camera parameter estimation. It should be noted that the great difference in image coverage between the tracker camera with the wide-angle lens and the two scene cameras does not leave much room for a variety of viewpoints towards the combined calibration object.

3.2.7 Real-Time Tracking

During the real-time operation of the system, we want to keep track of the head pose $\boldsymbol{H}_{\mathrm{tf}}$ and the pose $\boldsymbol{H}_{\mathrm{tp}}$ of hand-held tools for each incoming tracker camera image. The marker detection strategy provides us with the centers of n projected markers $\{\tilde{\boldsymbol{q}}_{\mathrm{i_t},\mu} \in \mathcal{P}^2 \mid \forall \mu \in M\}$, where $M = \{1, 2, \ldots, n\}$.

We use some simple heuristics to separate the markers of different sets in the camera image. Thus we have the image coordinates of the markers which frame the workspace $\{\tilde{\boldsymbol{q}}_{\mathrm{i_t},\mu} \in \mathcal{P}^2 \mid \forall \mu \in M_{\mathrm{f}}\}$ and the image coordinates of the markers which are attached to the hand-held tool $\{\tilde{\boldsymbol{q}}_{\mathrm{i_t},\mu} \in \mathcal{P}^2 \mid \forall \mu \in M_{\mathrm{p}}\}$, where $M_{\mathrm{f}} \subset M$ and $M_{\mathrm{p}} \subset M$. We designed the marker sets in a way that three characteristic markers can easily be distinguished from the rest of the set in the tracker camera image: The two top-most outstanding markers on the marker frame together with the marker on the oval frame that is central to both can easily be deduced from the 2D constellation of the extracted markers in the image. Similarly, the central marker of a marker cluster as well as the single large marker on the circle of markers around it together with its right-most neighbor define the three characteristic markers on a marker cluster. We apply the closed-form 3-point algorithm (i.e., described in [Har91]) to compute the initial estimate of the pose. The 3-point algorithm usually has multiple solutions, which can be narrowed down by taking the other extracted markers (of the same marker set) into consideration. The simplest approach is to choose the solution that minimizes the reprojection error for these markers in the image. We uniquely label all markers that belong to the marker frame (with the labeling function $\xi_{\mathrm{f}} : M_{\mathrm{f}} \mapsto M_{\mathrm{f}}$) and all markers that belong to the marker cluster of the tool or pointer (with the labeling function $\xi_{\mathrm{p}} : M_{\mathrm{p}} \mapsto M_{\mathrm{p}}$).

After extracting and labeling the coordinates of the projected markers, we can use the set of 2D–3D point correspondences $\{(\tilde{\boldsymbol{q}}_{\mathrm{i_t},\xi_{\mathrm{f}}(\mu)}, \tilde{\boldsymbol{q}}_{\mathrm{f},\mu}) \mid \forall\mu \in M_{\mathrm{f}}\}$ to estimate the pose $\boldsymbol{H}_{\mathrm{tf}}$ of the tracker camera and the set $\{(\tilde{\boldsymbol{q}}_{\mathrm{i_t},\xi_{\mathrm{p}}(\mu)}, \tilde{\boldsymbol{q}}_{\mathrm{p},\mu}) \mid \forall\mu \in M_{\mathrm{p}}\}$ to estimate the pose $\boldsymbol{H}_{\mathrm{tp}}$ of the hand-held tool. Here $\{\tilde{\boldsymbol{q}}_{\mathrm{f},\mu} \in \mathcal{P}^3 \mid \forall\mu \in M_{\mathrm{f}}\}$ are the homogeneous coordinates of the frame markers in the marker frame coordinate system and $\{\tilde{\boldsymbol{q}}_{\mathrm{p},\mu} \in \mathcal{P}^3 \mid \forall\mu \in M_{\mathrm{p}}\}$ are the homogeneous coordinates of the markers of the tool or pointer in the pointer coordinate system.

Since closed-form solutions for the pose estimation problem with 2D–3D point correspondences, like the 3-point algorithm, minimize an algebraic distance, which in general does not correspond to a meaningful geometric distance, like the reprojection error, it is important to have a strategy which increases the stability of the tracking result. A good overview about the different pose estimation approaches can be found in [Qua99]. If the number of matched markers for a tracked marker set is at least seven, we use Tsai's [Tsa87] closed-form external calibration procedure, which takes all marker correspondences into account, and thus gives a better initialization for the following non-linear iterative optimization procedure. Otherwise we directly use the result of the 3-point algorithm as initialization.

To adjust and refine the six dimensional parameter space that describes the pose $\boldsymbol{H}_{\mathrm{tf}}$ we minimize the reprojection error:

$$\epsilon^2 = \sum_{\mu \in M_{\mathrm{f}}} \left\| \tilde{\boldsymbol{q}}_{\mathrm{i_t},\xi_{\mathrm{f}}(\mu)} - \boldsymbol{\delta}\left(\boldsymbol{A}_{\mathrm{t}} \hat{\boldsymbol{H}}_{\mathrm{tf}}^{-1} \tilde{\boldsymbol{q}}_{\mathrm{f},\mu}, \{\kappa_{\mathrm{t1}}, \kappa_{\mathrm{t2}}, \kappa_{\mathrm{t3}}\} \right) \right\|$$

$$\text{with} \quad \boldsymbol{H}_{\mathrm{tf}} = \underset{\hat{\boldsymbol{H}}_{\mathrm{tf}}}{\operatorname{argmin}}(\epsilon^2) \, . \tag{3.15}$$

Radial distortion of an undistorted projected point $\tilde{\boldsymbol{p}}_{\mathrm{u}} \in \mathcal{P}^2$, caused by the tracker camera optics, is modeled according to (2.8):

$$\boldsymbol{\delta}(\tilde{\boldsymbol{p}}_{\mathrm{u}}, \{\kappa_{\mathrm{t1}}, \kappa_{\mathrm{t2}}, \kappa_{\mathrm{t3}}\}) = \tilde{\boldsymbol{p}}_{\mathrm{u}} \left(1 + \kappa_{\mathrm{t1}} \|\tilde{\boldsymbol{p}}_{\mathrm{u}}\|^2 + \kappa_{\mathrm{t2}} \|\tilde{\boldsymbol{p}}_{\mathrm{u}}\|^4 + \kappa_{\mathrm{t3}} \|\tilde{\boldsymbol{p}}_{\mathrm{u}}\|^6\right) . \tag{3.16}$$

We use the Levenberg-Marquart optimization to perform the minimization (3.15), where we choose conditions to stop the iteration within a maximum processing time of 5 milliseconds leading to sub-pixel overlay precision as will be shown in Chapter 4. A corresponding optimization is performed for $\boldsymbol{H}_{\mathrm{tp}}$.

Figure 3.10 illustrates the sequence of operations to calculate the pose of the tracker camera with respect to the marker frame. The same pose estimation scheme is applied to calculate the pose of hand-held tools, i.e., to estimate $\boldsymbol{H}_{\mathrm{tp}}$.

The resulting pose $\boldsymbol{H}_{\mathrm{tf}}$ of the tracker camera and pose $\boldsymbol{H}_{\mathrm{tp}}$ of the hand-held tool can be utilized by (3.10) to calculate where (in the medical data coordinate system) the user interacts

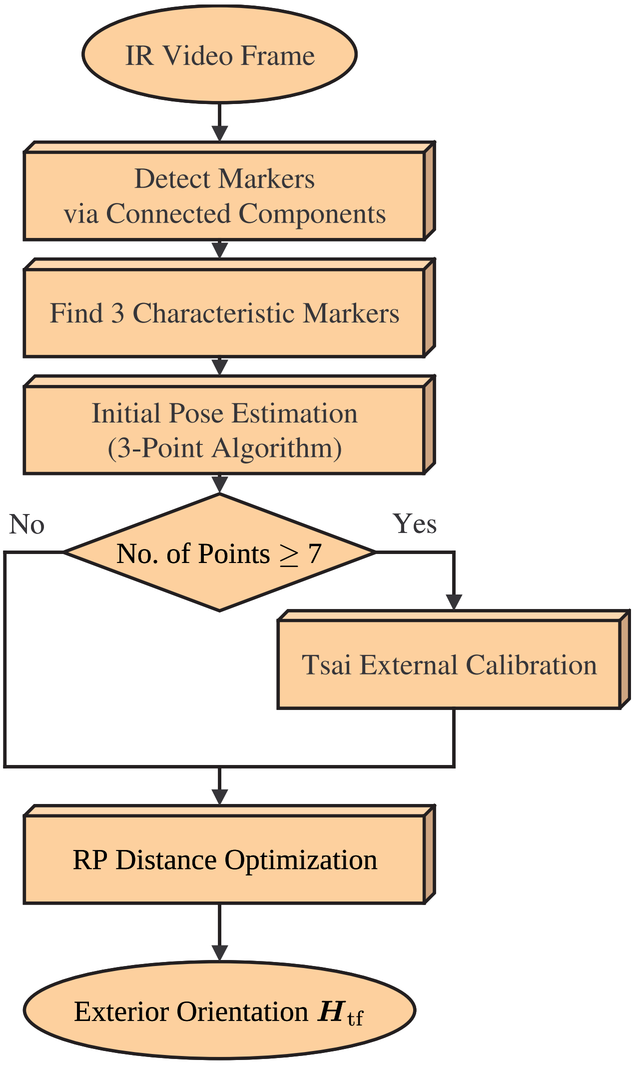


Figure 3.10: Sequence of operations for the estimation of the pose of the tracker camera during the real-time operation phase of the AR system, starting with the connected component analysis (Chapter 3.2.4). The non-linear optimization procedure (last step) to minimize the reprojection error is a crucial part to realize robust pose estimates from the incoming video sequence and thus to achieve jitter-free augmentation results.

with the medical dataset. Since all three cameras are genlocked, i.e., they all take their images synchronized at the same time, $\boldsymbol{H}_{\mathrm{tf}}$ can further be applied to calculate (3.8) and (3.9) for any point P whose coordinates are given in the medical data coordinate system. Thus (with equations analogous to (3.11) and (3.12)), we can deduce the coordinates of the projections of P in the left and right images of the scene cameras at that time.

For each incoming camera image, our tracking approach estimates the pose parameters without taking time continuity into account. On the one hand, we think that head and hand movements are highly unpredictable motions and thus do not consider approaches such as Kalman filtering for the pose parameters. On the other hand, as we will show in the following sections, with our current tracking approach we achieve stable and jitter-free video augmentations without any time-lag between video and graphics.

Chapter 4

RAMP—Technical Evaluation

The introduction of a unique and novel video see-through augmented reality system as described in the previous chapter, which exhibits real-time performance with 30 stereoscopic XGA images per second, necessitates a quantitative analysis and optimization of its main system properties. This includes the accuracy of the registration parameters (mainly determined by the calibration and tracking accuracy), its optimization of marker design parameters to achieve minimum reprojection errors and jitter of the graphics, as well as the actual perceived precision and value when utilized for guidance in user-driven tasks.

4.1 Calibration and Tracking

In Chapter 3.2.6 and Chapter 3.2.7 we described the unique approach to tracking with the RAMP system and its calibration routine for the camera triplet. We want to achieve stable and accurate graphics augmentations, which depend on precise calibration and tracking results.

4.1.1 Camera Calibration Accuracy

To quantitatively analyze the offline camera calibration procedure, we consider the residual reprojection error of the calibration markers. We utilized 10 calibration images for each camera from different view points around the calibration object. While aiming the optical axis centrally towards the calibration object we rotated the camera around towards the left, right, top, and bottom of the calibration object, up to an angle of 45 degrees with respect to the flat marker discs, to capture 10 representative view points. To utilize the retro-reflective capabilities of the markers also for the calibration of the scene cameras, we used a strong light source that we held just above the camera triplet pointing in the same direction as the scene cameras.

The calibration of the two scene cameras results in a mean residual reprojection error of about 0.22 pixels with a standard deviation of 0.15 pixels. The calibration of the tracker camera exhibits a mean residual reprojection error of about 0.16 pixels with a standard deviation of 0.09 pixels, which translates into a mean object space error of 0.27mm with a standard deviation of 0.15mm. The slight difference in calibration accuracy of the tracker camera in comparison to the scene cameras is most likely caused by the use of two different calibration objects that needed to be built for the two types of cameras capturing significantly different field-of-views (see Figure 3.8 and Figure 3.9). Furthermore, the infrared-pass filter of the tracker camera filters out the visible light of the background, making almost only the evenly-lit retro-reflective markers visible, whereas the calibration images of the scene cameras are influenced by ambient light causing slightly inhomogeneous marker and background brightness.

4.1.2 Tracking Accuracy

During the real-time tracking phase, we distinguish the pose estimation of the tracker camera with respect to the marker set that frames the workspace for head tracking (Chapter 3.2.2) from the pose estimation of a marker cluster for the tracking of hand-held tools (Chapter 3.2.3). The mean residual reprojection error of the marker coordinates of the frame marker set varies between 0.035 and 0.12 pixels in the tracker camera image, depending on the viewing direction of the tracker camera. A more angular view (up to 45 degrees), with respect to the marker frame oval, typically results in a slightly better pose estimation and hence a smaller reprojection error. This is caused by the better distribution of the markers in all 3 dimensions, especially in the direction of the optical axis, when the frame is viewed from an angle, such as from 45 degrees to the right or left side. However, in any case, this sub-pixel accuracy allows us to overlay graphics onto the video view of the scene cameras which seems to be perfectly aligned with real objects in the scene and which does not exhibit any visible jitter.

Especially the last tracking step, the non-linear optimization of the reprojection error (see Figure 3.10), justifies its importance when we look at the reprojection error before the optimization. After the 3-Point algorithm the mean residual reprojection error of the frame marker coordinates is between 0.065 and 3.25 pixels. After the non-linear optimization this reprojection error reduces to the afore mentioned range of 0.035 to 0.12 pixels. It should also be remarked that a small reprojection error in the tracker camera, in the sub-pixel range, translates into a much bigger error in the scene cameras, due to the use of a wide-angle lens. Therefore, it is mainly because of the non-linear optimization routine that we achieve jitter-free augmentations.

In many of our AR experiments we used a marker cluster to track the pose of a biopsy or an ablation needle. Since the needle tip is relatively far away from the cluster of markers (see Figure 3.5), it was important to design a small marker cluster in a way that the estimated position of the needle tip in the tracker camera coordinate system exhibits only a minimal jitter. After optimizing the design of this marker cluster (see Chapter 4.2), the jitter of the estimated needle tip position is less than 1mm, which results in a stable augmentation of the video view with graphics that represents the needle.

4.2 Marker Cluster Optimization

Because of the size of the marker set and its large coverage in the tracker camera image, we achieved stable augmentations of stationary objects that are framed by the surrounding marker set. No swimming or jitter of the graphics overlay is visible. Nevertheless, augmenting hand-held tools that are tracked by means of an attached marker cluster resulted originally in graphics overlays that were visibly jittering on top of the video view. Our first marker cluster was coplanar and we realized that a thorough analysis was necessary to optimize such a marker cluster design to eliminate visible jitter but yet keep the marker cluster itself small for practical reasons [Vog02].

4.2.1 Marker Cluster Tracking Error

The outcome of the pose estimation algorithm (external camera calibration) is a 6D quantity encoding the translation and rotation. From the error of the 6D pose estimation we can deduce the 3D position error of a target point in the marker frame or cluster coordinate system in two forms. The first is a *biased (systematic) error*, which is the deviation of the mean of the 3D position from the true value. The second is a *jitter*, which is the variation of the 3D position in time around the mean position.

We assume that the biased pose error is related to the internal calibration accuracy, the precision of the 3D coordinates of the marker set, any bias that exists in the estimates of the 2D locations of the projected markers in the tracker camera image, and any biased error in the 3D–2D matching (pose estimation) algorithm. Also, because of the non-linear nature of the projection, any unbiased noise in the pose values reflects a biased error in the 3D position.

Studying the biased error in the 3D position of a target point is a challenging subject. This is mainly because the dimension of the parameter space for the problem is quite high. In this study, our primary goal is to achieve a jitter-free and stable overlay in an AR setup. Therefore, we only consider the jitter error, which is dominantly caused by the error of the estimation of the 2D marker locations in the tracker camera image. By characterizing this error, we are able to estimate the resulting error of the pose and consequently predict the amount of jitter in the 3D position of any target point. We assume that the biased error is negligible mainly under the assumption that the parameters affecting the biased error are pre-calibrated offline and can be acquired with high degrees of accuracy.

To compare merits of various marker cluster configurations, we consider two forms of error measures:

- the 6D pose estimation error (jitter of calculated rotation and translation)

$$\boldsymbol{\epsilon_6} = \left(\left\|\bar{\theta}_{\mathrm{x}} - \hat{\theta}_{\mathrm{x}}\right\| \; \left\|\bar{\theta}_{\mathrm{y}} - \hat{\theta}_{\mathrm{y}}\right\| \; \left\|\bar{\theta}_{\mathrm{z}} - \hat{\theta}_{\mathrm{z}}\right\| \; \left\|\bar{t}_{\mathrm{x}} - \hat{t}_{\mathrm{x}}\right\| \; \left\|\bar{t}_{\mathrm{y}} - \hat{t}_{\mathrm{y}}\right\| \; \left\|\bar{t}_{\mathrm{z}} - \hat{t}_{\mathrm{z}}\right\|\right)^{\mathrm{T}} \,, \tag{4.1}$$

- the 3D error of a target point within the desired workspace of the marker cluster

$$\boldsymbol{\epsilon_3} = \left(\|\bar{p}_{\mathrm{cx}} - \hat{p}_{\mathrm{cx}}\| \; \|\bar{p}_{\mathrm{cy}} - \hat{p}_{\mathrm{cy}}\| \; \|\bar{p}_{\mathrm{cz}} - \hat{p}_{\mathrm{cz}}\|\right)^{\mathrm{T}} \,, \tag{4.2}$$

where $\hat{}$ denotes the current estimate and $\bar{}$ the mean estimate. The 6D pose is parameterized by three sequential rotations $\boldsymbol{R}_{\mathrm{x}}(\theta_{\mathrm{x}}), \boldsymbol{R}_{\mathrm{y}}(\theta_{\mathrm{y}}), \boldsymbol{R}_{\mathrm{z}}(\theta_{\mathrm{z}}) \in \mathbb{R}^{3\times3}$ around the three axes of the coordinate system with rotation angles $\theta_{\mathrm{x}}, \theta_{\mathrm{y}}, \theta_{\mathrm{z}} \in \mathbb{R}$ and a translation along the three coordinate axes with $t_{\mathrm{x}}, t_{\mathrm{y}}, t_{\mathrm{z}} \in \mathbb{R}$. Correspondingly, a point P with coordinates $\boldsymbol{p}_{\mathrm{p}} = (p_{\mathrm{px}} \; p_{\mathrm{py}} \; p_{\mathrm{pz}})^{\mathrm{T}} \in \mathbb{R}^3$ in the marker cluster coordinate system is transformed to camera coordinates $\boldsymbol{p}_{\mathrm{c}} = (p_{\mathrm{cx}} \; p_{\mathrm{cy}} \; p_{\mathrm{cz}})^{\mathrm{T}} \in \mathbb{R}^3$ by

$$\boldsymbol{p}_{\mathrm{c}} = \boldsymbol{R}_{\mathrm{z}}(\theta_{\mathrm{z}})\boldsymbol{R}_{\mathrm{y}}(\theta_{\mathrm{y}})\boldsymbol{R}_{\mathrm{x}}(\theta_{\mathrm{x}})\boldsymbol{p}_{\mathrm{p}} + (t_{\mathrm{x}} \; t_{\mathrm{y}} \; t_{\mathrm{z}})^{\mathrm{T}} \,.$$

The first error measure is the most comprehensive one for error evaluation, whereas the second form only focuses on the errors in the target point(s) and thus is intuitively more comprehensible. For instance, an error distribution in the position of a target point can readily be presented by a 95% confidence ellipsoid. It is also noteworthy to mention that the overlay jitter of the projected target point onto an arbitrary plane can be extracted using the 3D error distribution of that point.

We first identify the variation of the extracted 2D locations of the centers of the projected markers in the camera image and then analyze the error measures discussed above for different cluster configurations using Monte Carlo simulations and by performing real experiments.

4.2.2 Evaluation of the Cluster Design

Single camera tracking of hand-held tools by means of (extrinsic) camera calibration becomes possible by attaching marker clusters as described in Chapter 3.2.3. Given the exact 3D locations of the markers in their own coordinate system and the 2D locations of the marker projections in the tracker camera image as input for the calibration technique (Chapter 3.2.7) leads to the rigid body transformation between the marker cluster coordinate system and the camera coordinate system, i.e., the 6D pose $\boldsymbol{H}_{\mathrm{tp}}$ of the marker cluster. Thus, three groups of input parameters influence the result of the pose estimation:

- the intrinsic parameters of the tracker camera,
- the 3D coordinates of the markers in the cluster coordinate system, and
- the 2D coordinates of the projections of the markers in the tracker camera image.

On the one hand, it is necessary to precisely determine these input parameters to achieve a high accuracy for the pose estimation. Assuming fixed intrinsic camera parameters and rigidly mounted markers, only the last item, i.e., the extraction of the 2D coordinates from the camera image, becomes a crucial factor during real-time tracking of the marker cluster. Pixel jitter in the camera image therefore directly affects the stability of the estimated 6D pose. To what extent this 2D jitter translates into a jitter of the 6D pose of the marker cluster (and the attached tool) depends, on the other hand, also on the design of this cluster.

Analyzing the influence of the structure of the marker cluster on the stability of its estimated 6D pose is the major interest in this study. The following illustrates how we evaluated a family of clusters in computer simulations and experimental measurements in terms of the error measures discussed in Chapter 4.2.1.

At first we measure and characterize the jitter of the 2D locations of the projected markers in the real image of the tracking camera for different marker sizes and positions. Assuming an unbiased Gaussian distributed error around the true 2D marker locations, we then use this information to investigate the outcome of the pose estimation based on a simulated 2D projection of a variety of synthesized 3D marker clusters. This is done by modeling 3D marker clusters with the computer and projecting the 3D marker centers under certain poses into the image plane by using the camera model of our real tracking camera. After adding synthetic noise, which is defined through the prior real observations, to the 2D data of the projected marker centers, we estimate the 6D pose of the marker cluster with the calibration method that is used in our real tracking system. For more intuitive error visualizations, we represent the 6D pose $\boldsymbol{H}_{\mathrm{tp}}$ by three

Cardan angles ($\theta_\mathrm{x}, \theta_\mathrm{y}, \theta_\mathrm{z}$) and a translation vector $\boldsymbol{t} = (t_\mathrm{x}\ t_\mathrm{y}\ t_\mathrm{z})^\mathrm{T}$ as previously noted:

$$\boldsymbol{p}_\mathrm{c} = \boldsymbol{R}\boldsymbol{p}_\mathrm{p} + \boldsymbol{t} \quad \text{with} \quad \boldsymbol{R} = \boldsymbol{R}_\mathrm{z}(\theta_\mathrm{z})\boldsymbol{R}_\mathrm{y}(\theta_\mathrm{y})\boldsymbol{R}_\mathrm{x}(\theta_\mathrm{x}) \quad . \tag{4.3}$$

Thereby, $\boldsymbol{p}_\mathrm{p} \in \mathbb{R}^3$ and $\boldsymbol{p}_\mathrm{c} \in \mathbb{R}^3$ stand for the 3D coordinates of a point in the marker cluster coordinate system and the camera coordinate system, respectively. The z-axis of the camera coordinate system is aligned with the optical axis of the camera and points into the scene. For the marker cluster coordinate system we chose the one introduced in Figure 3.5.

We quantify and analyze the error measures of Chapter 4.2.1 for a variation of the marker cluster parameters and a variation of different poses. Additionally, we also build two marker clusters—one coplanar and one non-coplanar—and measure the deviations in the pose estimation for each of them in an experimental setup.

4.2.3 2D Single Marker Jitter

In this first stage, we use single markers of different sizes and analyze the jitter of the extracted 2D locations in the image of the tracker camera. The method for extracting the center of a projected marker utilizes all the available image information by calculating the moments on the gray-values of the retroreflective markers as described in Chapter 3.2.4.

We choose a setup where several markers in a stationary pose are projected into the image of the tracking camera. The range of the bounding box size is 80 to 1200 square pixels for the projected markers. After taking several thousand measurements for a variety of marker sizes, we conclude that the noise in the extracted 2D coordinates of the centers of the projected markers decreases insignificantly with an increased size of the projected marker in the image. By calculating the covariance matrix on all extracted 2D coordinates we can furthermore deduce that the correlation factor between the noise in the horizontal and vertical coordinate is about 0.4 to 0.6. The noise for different markers of a marker cluster is also correlated with a correlation factor of 0.4 to 0.6. The standard deviation of the noise of the horizontal and vertical coordinate consistently varies from 0.013 to 0.010 pixel for a bounding box size of 80 to 1200 square pixels.

We assume that the measured 2D jitter is caused by the tracking hardware, i.e., mainly the video digitizer board and the tracking camera itself, which also causes the correlation between the noise in the x- and y- direction of the image. Consequently, a jitter in the gray values causes a jitter of the calculated 2D marker centers in the camera image. The tests show that the image size of the projected marker seems not to be an important factor as long as the marker spans over several pixel rows and columns in the camera image.

Attribute	*Series 1*	*Series 2*	*Series 3*	*Series 4*	*Series 5*
Cluster height	0. . . 45 mm	30 mm	30 mm	0 mm	30 mm
Cluster radius	45 mm	15. . . 60 mm	45 mm	45 mm	45 mm
Number of markers	8	8	7. . . 16	8	8
Cluster pose	fixed	fixed	fixed	variable	variable

Table 4.1: Attributes of the studied marker clusters.

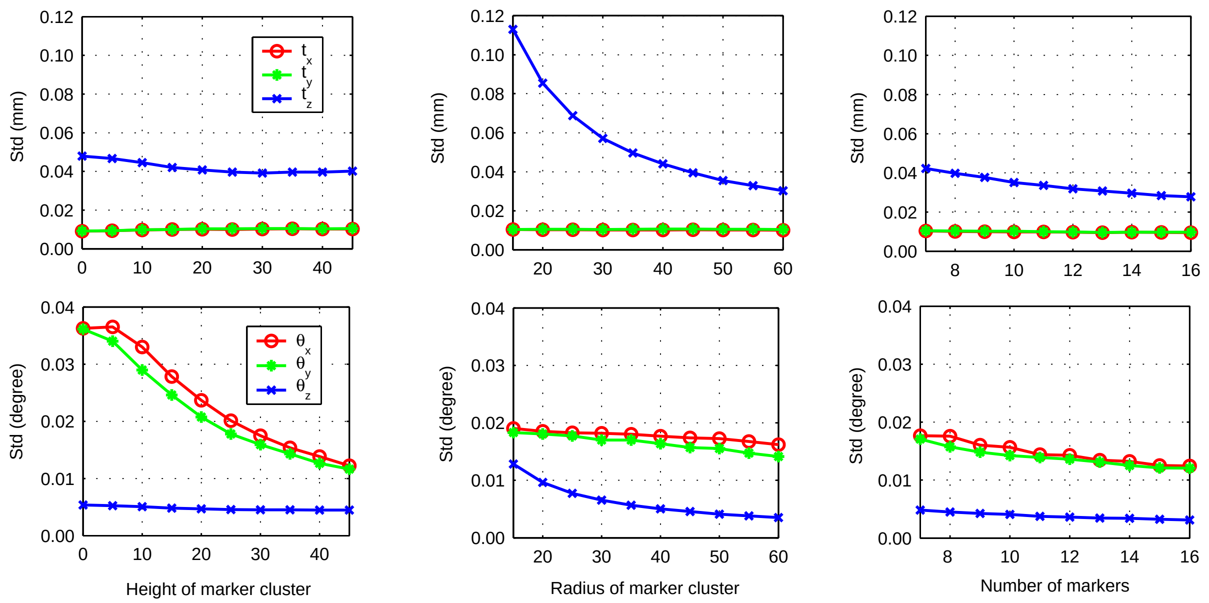


Figure 4.1: Simulation error plots for the estimated pose, *(top)* translation and *(bottom)* rotation, for changes in the characteristics of the synthesized marker cluster: *(left)* height, *(middle)* radius, and *(right)* number of markers.

4.2.4 Simulated Variation of the Cluster Parameters

To study the sensitivity to noise using Monte Carlo simulations, we synthesize 3D marker clusters with different radiuses, heights, and numbers of markers. At each series of experiments one of the three parameters of the marker cluster is changed while the other two are kept constant. Table 4.1 shows the chosen values (Series 1 to 3). The internal parameters of our real tracker camera are known through a previous calibration procedure and used to model the camera in our simulations. The external pose is the same for each investigated marker cluster, with $\theta_{\mathrm{x}} = 10^\circ$, $\theta_{\mathrm{y}} = 5^\circ$, $\theta_{\mathrm{z}} = 0^\circ$, $t_{\mathrm{x}} = 0\,\mathrm{mm}$, $t_{\mathrm{y}} = 5\,\mathrm{mm}$, $t_{\mathrm{z}} = 500\,\mathrm{mm}$. It is a typical configuration when a marker cluster is attached to a hand-held tool.

This first set of experiments is performed to observe the effect of each parameter of a marker cluster on the overall accuracy. It includes the mentioned three scenarios where in each of them

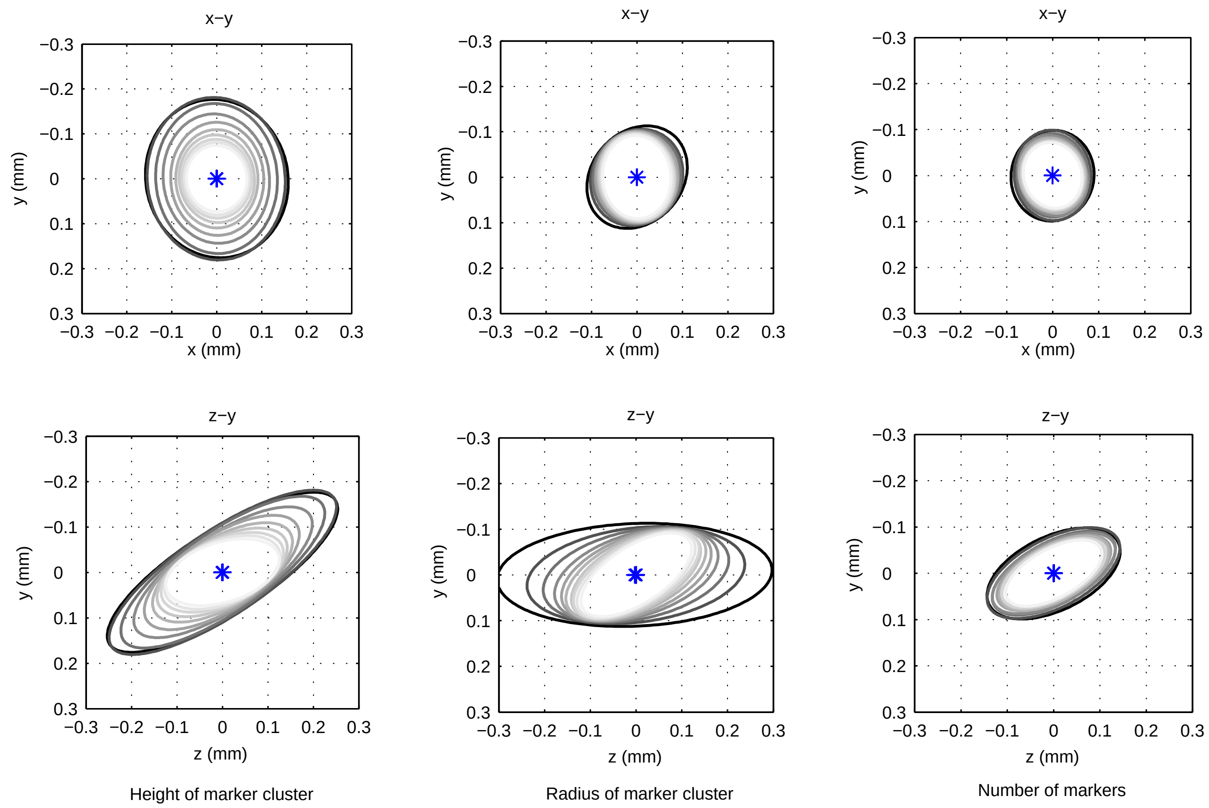


Figure 4.2: Projection of the 3D 95% confidence ellipsoid for a target point ([-100 mm, -100 mm, -100 mm] in the marker cluster coordinate system) into the *(top)* x-y plane and *(bottom)* y-z plane, for changes in the characteristics of the synthesized marker cluster: *(left)* height, *(middle)* radius, and *(right)* number of markers. The brightness of the ellipse corresponds to the changing variable along the horizontal coordinate axis in Figure 4.1.

only one parameter changes. After projecting the synthesized 3D marker cluster into the image plane of the modeled tracker camera, we add a high number (i.e., 10000) of samples of noise to the 2D locations of the simulated projections of the marker centers based on the experimentally determined noise characteristics (Chapter 4.2.3).

The measured variation of the estimated pose is depicted in Figure 4.1. The top row plots in Figure 4.1 show the translational variation in millimeter and the bottom row plots the rotational variations (Cardan angles) in degree. The three columns in Figure 4.1 represent the results of the three test series with marker cluster parameters according to Table 4.1 (Series 1 to 3).

Figure 4.2 illustrates how the variation in pose in Figure 4.1 translates into a variation of the 3D coordinate of a target point in the coordinate system of the tracker camera. The target point is chosen to be 173.2 mm away from the center of the marker cluster coordinate system at [-100

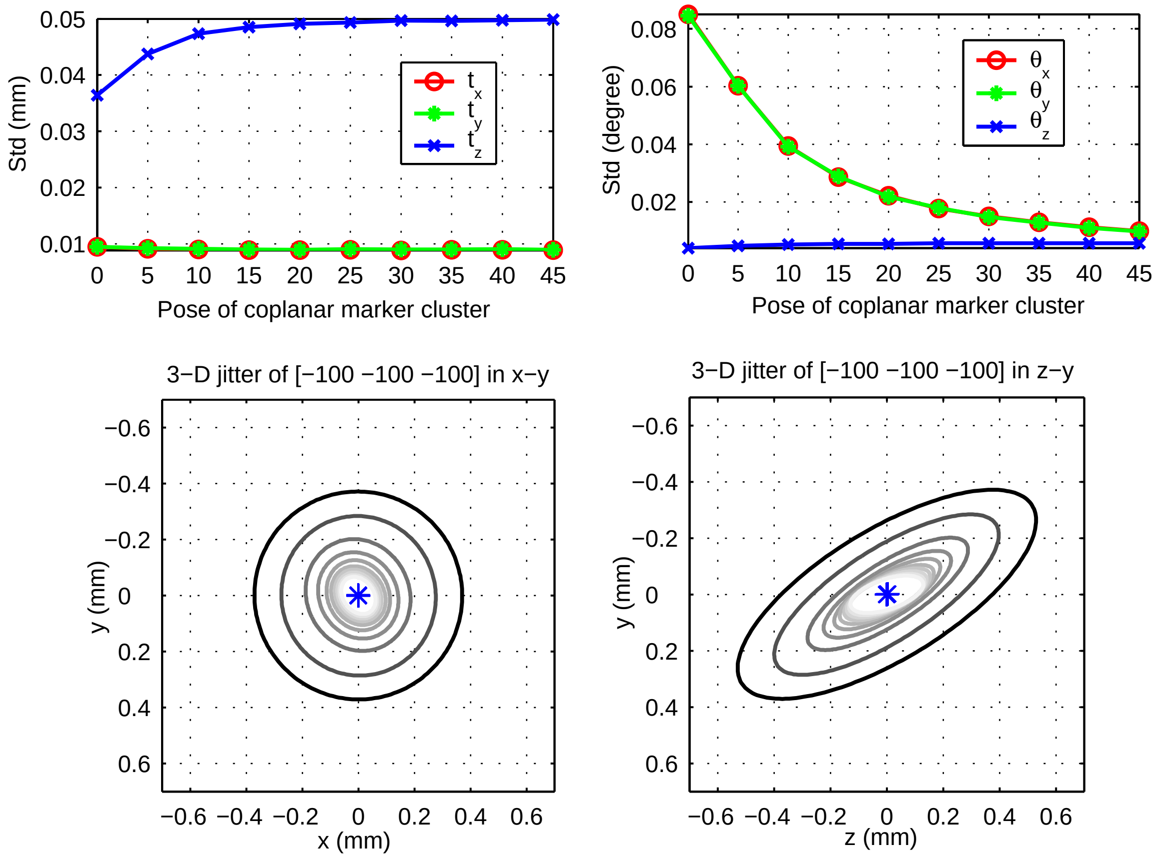


Figure 4.3: Simulation error plots for the estimated pose of a coplanar marker cluster for various poses: *(top left)* translation error and *(top right)* rotation error. The horizontal coordinate axes show the tilt of the circle plane of the marker cluster against the tracking camera plane in degree. The bottom figures show the projection of the 3D 95% confidence ellipsoid for a target point ([-100 mm, -100 mm, -100 mm] in the marker cluster coordinate system) into the *(bottom left)* x-y plane and *(bottom right)* y-z plane.

mm, -100 mm, -100 mm]. Figure 4.2 contains three columns for the parameter variations. The top row illustrates the projection of a 95% confidence ellipsoid into the x-y plane and the bottom row shows the projection of the same ellipsoid into the y-z plane of the tracker camera coordinate system assuming a right handed system with the z-axis as the optical axis pointing into the scene. Different ellipses with various shades of gray indicate the variation of the parameter of the marker cluster, whereby brighter ellipses correspond to a higher parameter value.

Referring to Figure 4.1, it is obvious that the height of the marker cluster, and as such the distribution of the markers in all three dimensions, is crucial for the accuracy of the estimated tilt of the x-y plane of the camera coordinate system against the x-y plane of the marker cluster coordinate system. The standard deviations of the errors in the rotation angles θ_x and θ_y significantly

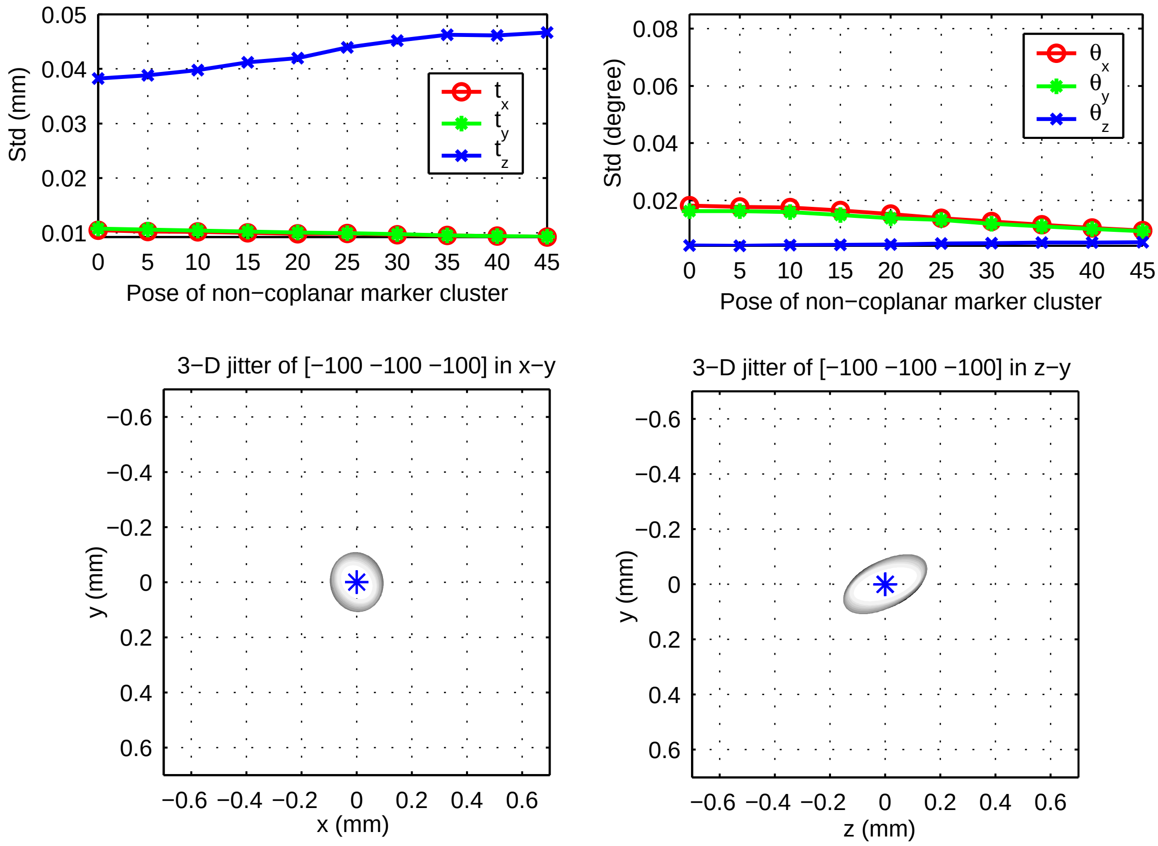


Figure 4.4: Simulation error plots for non-coplanar marker cluster. Description as in Figure 4.3.

decrease when the markers of the cluster are better distributed in 3D space instead of being coplanar. Figure 4.1 further reveals the fact that the jitter of the center of the marker cluster (represented by the translation vector $\boldsymbol{t}$) is mainly along the line-of-sight of the tracker camera. This error can be reduced by using larger marker clusters.

4.2.5 Simulated Variation of the Cluster Pose

The second set of simulations uses two specific marker clusters, one coplanar and one non-coplanar, to investigate how the errors depend on the cluster pose. Series 4 and 5 in Table 4.1 refer to these marker cluster configurations, respectively. We simulate a variation in the pitch angles (θ_x) from 0°to 45°. Starting in a parallel configuration of the camera plane and the plane of the marker cluster circle we stepwise tilt the marker cluster—keeping the center at the same location—until the planes reach a 45°angle to each other. Then, we repeat the same procedure as explained in Chapter 4.2.4 to determine the error plots for the pose and consequently the 3D

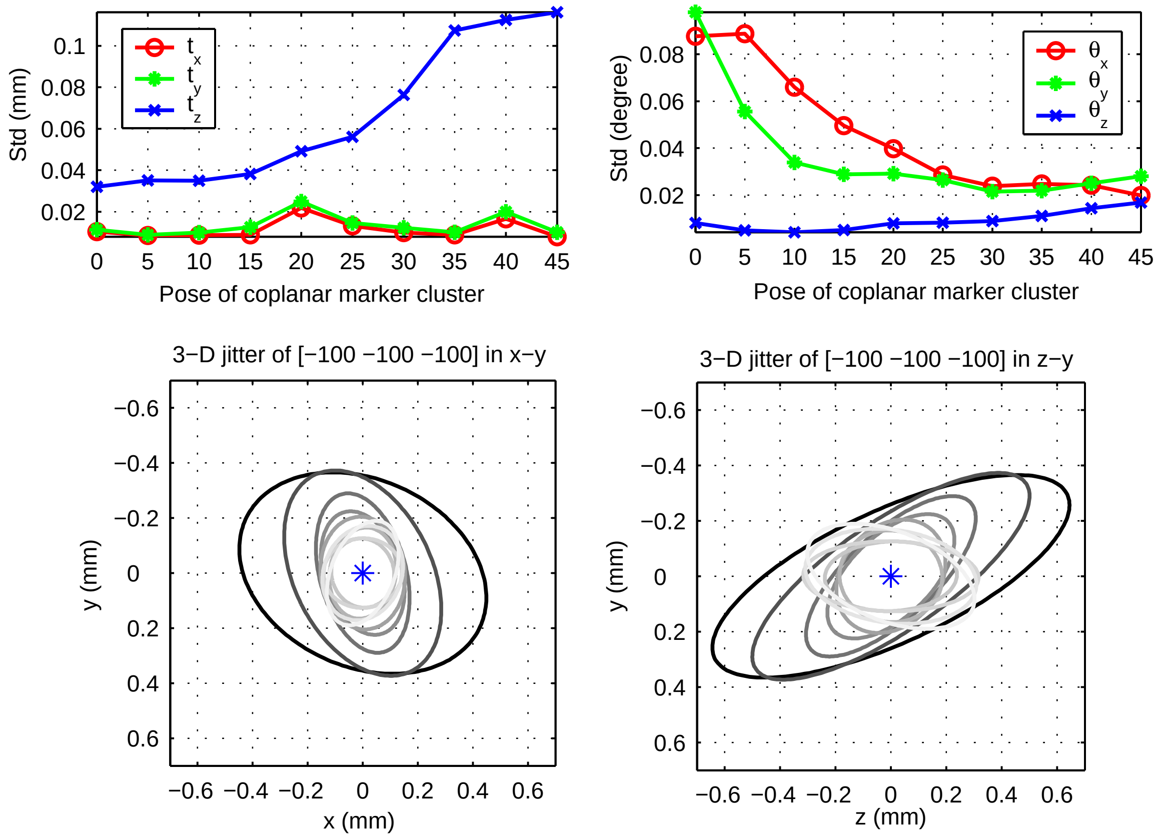


Figure 4.5: Error plots for real coplanar marker cluster. Description as in Figure 4.3.

95% confidence ellipsoid for the target point ([-100 mm, -100 mm, -100 mm] in the marker cluster coordinate system). These sets of plots are depicted in Figure 4.3 and Figure 4.4 for the synthesized coplanar and non-coplanar marker cluster, respectively.

The graphs show very clearly how the jitter in the rotation angles θ_x and θ_y is affected by the actual pose of the coplanar marker cluster. Especially a (near) top view at the cluster, which is a very common configuration in a real setup, reveals an intense source for jitter in these rotation angles. This further translates to a jitter of more than one millimeter of the chosen target point, caused by the leverage effect of the error in the rigid body transformation. Choosing a non-coplanar marker cluster can greatly improve the accuracy of θ_x and θ_y. The jitter of the target point reduces to about 0.3 mm.

These results basically confirm that the worst case for a coplanar marker cluster occurs, in terms of accuracy, when it is viewed from the top. In general, a non-coplanar cluster delivers a much more consistent error distribution that is comparable with the best case scenario of a coplanar cluster.

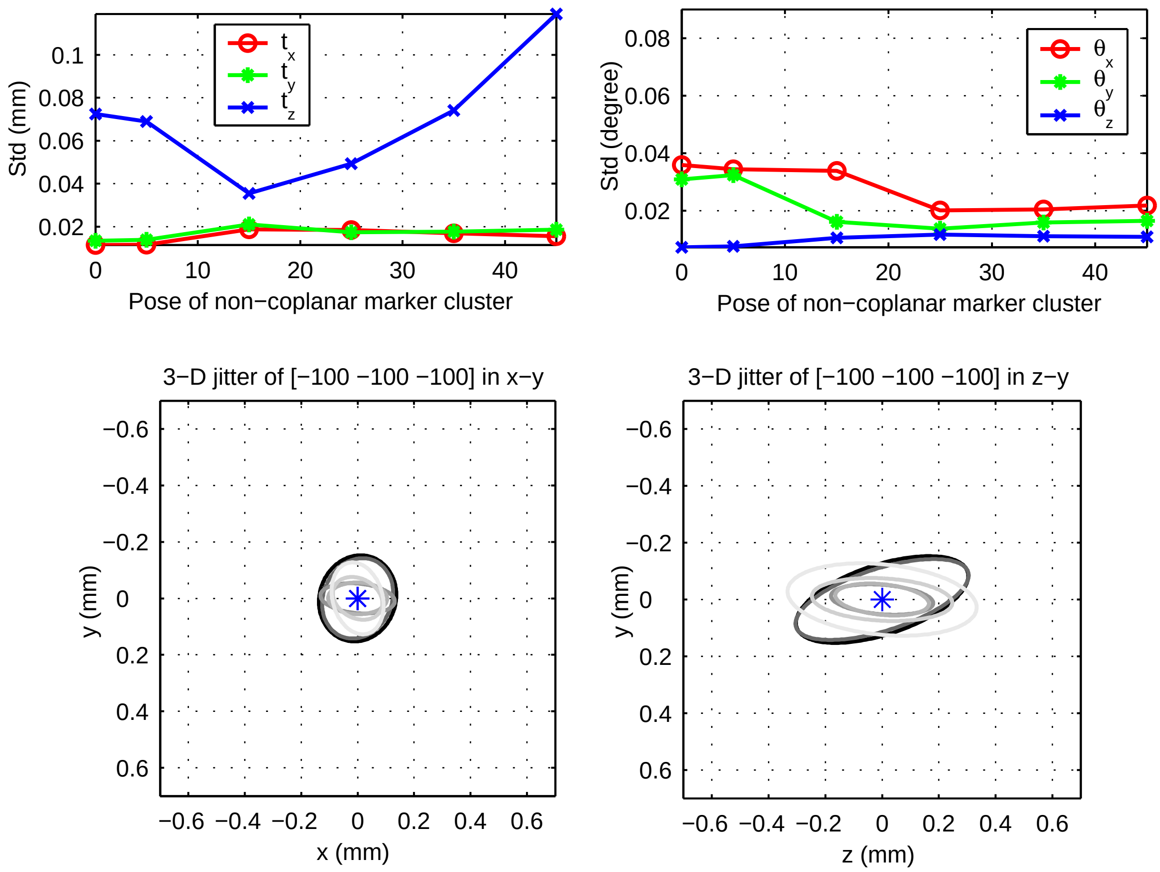


Figure 4.6: Error plots for real non-coplanar marker cluster. Description as in Figure 4.3.

4.2.6 Experimental Variation of the Cluster Pose

To verify the results of the computer modeled marker clusters we build two marker clusters, one coplanar and one non-coplanar. Their design is similar to the ones used in the simulations in the previous section. We use these two marker sets for tracking and record the estimated pose of our real-time tracker for 1000 samples while the marker cluster is stationary fixed. We repeat this test for different poses by varying the angle between the x-y plane of the marker cluster and the x-y plane of the tracking camera from 0˚to 45˚. The resulting jitter of the pose and consequently the projection of the 95% confidence ellipsoid of the target point (as chosen in the previous sections) are depicted in Figure 4.5 and Figure 4.6 for the coplanar and the non-coplanar marker cluster, respectively.

The similarity of these error graphs with the ones of the simulation in Chapter 4.2.5 validates the choice of the noise model and confirms the correctness of the study done for the marker configuration design.

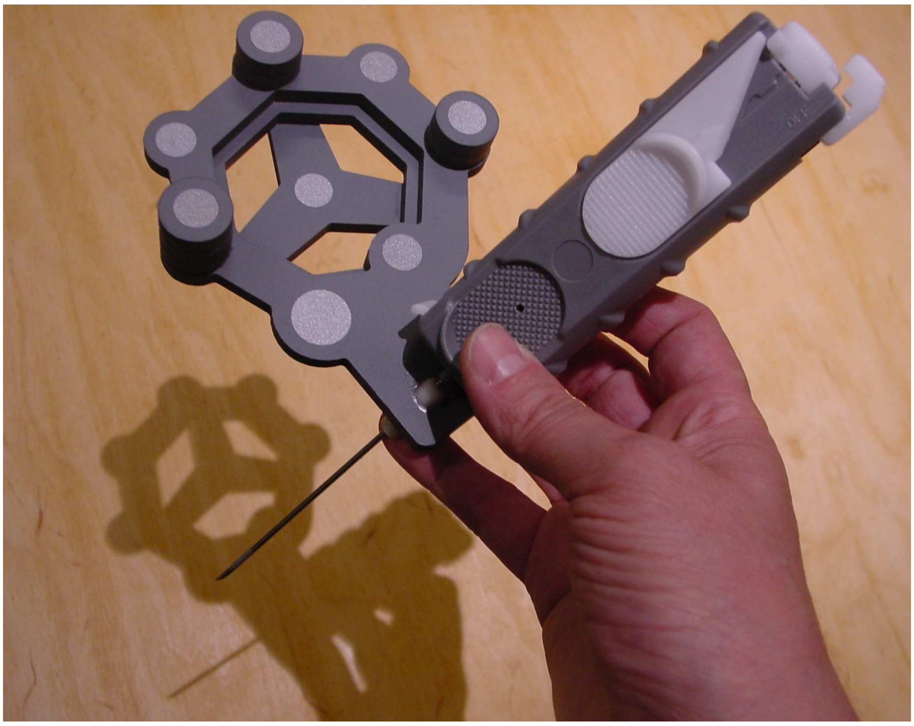


Figure 4.7: Biopsy needle with attached marker cluster for pose tracking. It should be noted that the displayed setup is designed for right-handed usage, since the marker cluster is attached tilted downwards, in an angle that provides best coverage by the tracker camera when held in the right hand. A similar cluster has been designed for left-handed usage.

4.3 AR Guided Navigation

The capability of tracking the head pose of the user as well as the pose of hand-held instruments provides the means to utilize our RAMP system for AR guided navigation tasks. Attaching a well designed marker cluster to an instrument such as a biopsy needle (see Figure 4.7) can provide 3-dimensional visualizations of the needle trajectory and potentially ensure its success for hitting a hidden target structure. We prepared and reported on a study [Sau02c] to evaluate the precision of such an AR guided navigation task to prove its effectiveness in a first laboratory setting.

4.3.1 Needle Placement Phantom

We designed a box with a set of hidden mechanical push-buttons to be used for a needle placement study (Figure 4.8). The push-buttons, which are the hidden targets to be touched with the needle tip, are small and cylindrical with a top diameter of 6mm. When they are being pushed down they depress the keys of an underlying USB keypad. If the user pushes down one of the

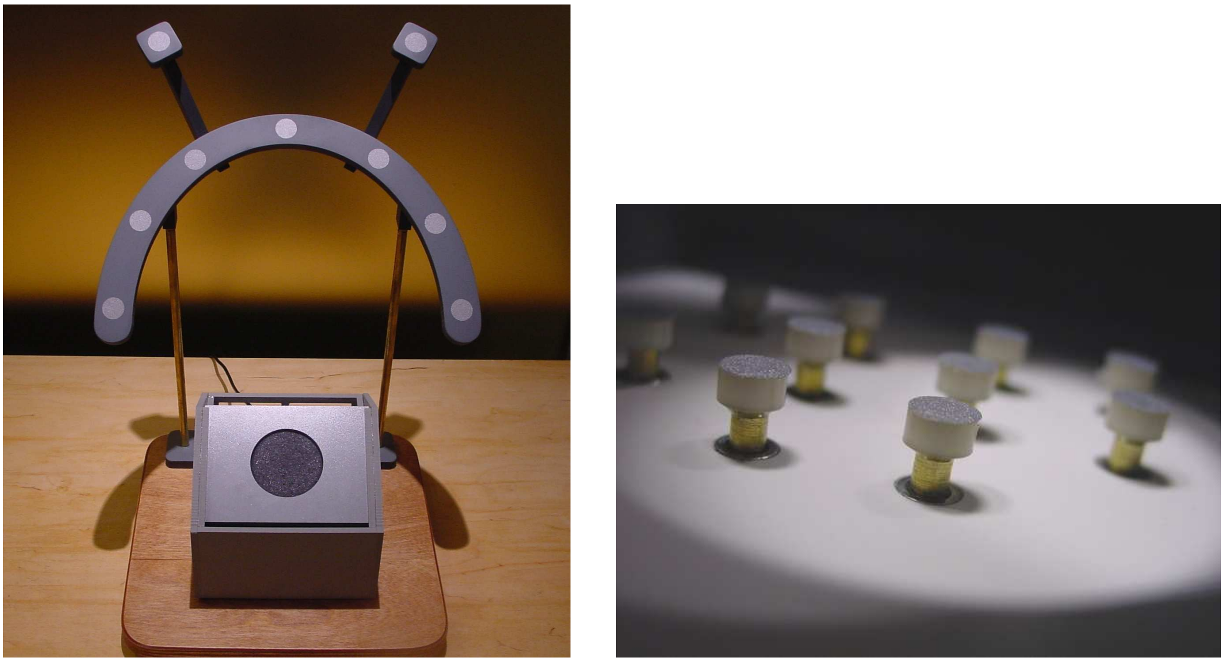


Figure 4.8: Phantom to evaluate the accuracy of AR guidance in a needle placement task. *Left:* Phantom box with a foam window that hides the underlying targets, which is framed by a set of retroreflective markers. *Right:* Cylindrical push-button targets with a head diameter of 6mm, 7cm underneath the foam surface of the phantom box.

targets with the needle, the computer provides acoustic feedback triggered by the USB-keypad connection. The slanted top face of the box provides a round window to access the targets with the needle. A thick foam pad is placed into this round window to hide the targets from the user's direct view. The 5cm thickness of the foam pad also provides mechanical resistance to the needle insertion. The distance between the top surface of the foam pad and the targets underneath is about 7cm.

The box, as depicted in Figure 4.8, is sitting on a platform which is framed by a marker set of seven coplanar markers on a half circle and two additional markers sticking out on little posts. Furthermore, we put retroreflective material onto the round heads of the push-button targets to easily be able to acquire the location of all the targets with respect to the marker frame coordinate system in a calibration step.

The head-mounted tracker camera, which tracks the marker frame, is also responsible for tracking the marker cluster that we attach to the biopsy needle as shown in Figure 4.7. The camera can reliably locate the individual markers while tilting the marker body within an angle range of about 45˚from the normal direction, where the marker cluster directly faces the camera.

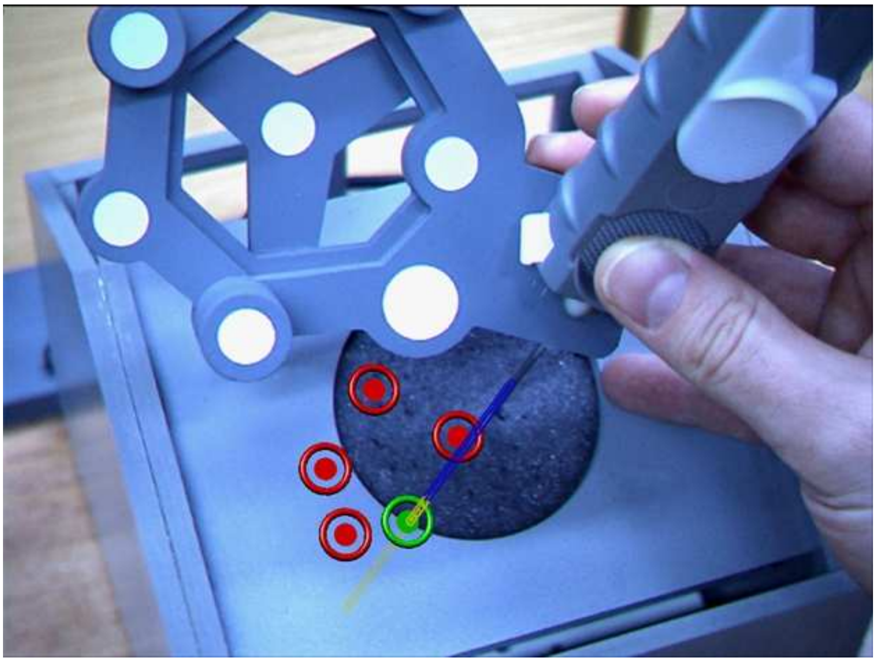


Figure 4.9: Augmented view for needle guidance. The tracked needle is visualized as a blue wireframe cylinder underneath the foam surface. The needle trajectory is represented as a yellow wireframe cylinder extending from the tip of the needle and guiding the user for proper targeting. The discs represent the targets 7cm below the foam surface. It should be noted that the stereoscopic and dynamic view of the AR system provides the user with intuitive spatial relationships that cannot be presented in this single image.

4.3.2 AR Visualization

The top surface of the hidden targets are visualized as flat discs to provide the user an augmented view of the phantom box. Each of these virtual target discs is surrounded by a ring, in shape of a rendered and shaded torus. This torus helps with the 3D perception of the location of the target. We chose a red color for this disc-torus target structure and have the computer switch to green when the needle is pointing toward the target. The needle itself is visualized as a blue wireframe cylinder. The extrapolation of the needle path is marked by a yellow wireframe cylinder. The intersection of the target discs and the extrapolated needle cylinder allows the user to judge whether the needle is pointing precisely toward the target. A typical augmented view of this setup that guides the user is shown in Figure 4.9. The needle is partially inserted through the foam window, positioned about 1cm above and correctly pointing to one of five targets shown.

The user perceives a very intuitively accessible augmented reality view, since our AR video system is running at the full standard video rate of 30 frames per second. Any time lag between the real and the virtual objects is eliminated because of the synchronized video and graphics. The virtual objects do not lag behind, neither do they seem to swim or jitter with respect to the real scene. As the augmented view shows the graphics firmly anchored in the real scene, the user can assess the information in a comfortable way. Overall, there is a time delay of about 0.1 seconds between an actual event and its display to the user.

4.3.3 User Study and Success-Rate

More than one hundred users have tested the AR guided needle experiment. Before we hide the real targets with the foam window we usually slide the cover aside to show them to the user. We also explain the needle visualization so that the user understands how to judge correct orientation and insertion depth of the needle. Initially, three targets are being shown. Every user is asked to perform three AR guided needle placements. When the user correctly hits one of the targets, by depressing the real piston with the needle tip, the computer plays an audio signal and the corresponding virtual target fades away.

Consistently, all users were able to hit the targets. In fact, the visualization is so intuitive and the visual feedback so conclusive, that one is basically not able to miss once one understands the basic concept. We observed that almost all users grasped the concept immediately, a few after a bit of experimentation with a learning curve below a minute. If time was said to be of importance most users could successfully and repeatedly perform the three needle placements at a rate of one per second, after only a few training trials.

To achieve stable needle tracking it is of importance to hold the needle in a way that the markers face toward the head-mounted tracker camera. This was the most common initial problem the test users had, but they quickly understood its correct handling. In our opinion, the fact that we can track the marker body only over about 45°away from the normal does not really represent a practical limitation for the needle placement. As long as the user is aware that he cannot turn the needle around its axis away from the tracker camera, the RAMP system provides stable, jitter-free AR visualization.

Since this user interface was experienced as very intuitive, Augmented Reality guidance may be especially helpful in an interventional setting when the user encounters complex hidden anatomy with a penetrating instrument, where vital structures like nerves or blood vessels have to be avoided while the instrument is advanced toward a target like a lesion. The study proved that the RAMP system can provide AR guidance with 100% success-rate in placing a needle into 6mm objects that are hidden 7cm below the surface. It not only gives intuitive access to understanding the 3D geometry of the anatomy, it also provides a comfortable and believable augmented reality experience, where the graphical structures appear firmly anchored in the video.

4.4 Comparison with other Navigation Methods

In Chapter 4.3 we studied and proved that augmented reality visualization with the RAMP system can successfully guide manual tasks with tracked instruments, such as placing a needle into a hidden target. But how does this type of AR guidance compare to other image-based navigation methods? In [Aza04] we report on a study that analyzes and compares user performance in needle placement procedures under four different image-based navigation approaches, including AR guidance.

4.4.1 Experimental Setup

The success of a needle placement procedure depends on reaching as close as possible to the center of the lesion, avoiding critical structures (vessels, etc.) in the surrounding of the lesion, and defining a straight needle path, to avoid needle bending. To analyze how different navigation methods influence the outcome of a needle placement, we created a setup that could be used for all methods. We modified the phantom box (Chapter 4.3.1) by removing the push-buttons in a way that it is mainly filled with foam to simulate real tissue during the needle insertion. We furthermore simulate the target lesions and "critical structures" in form of computer-generated spheres and bended tubes that are virtually located inside the foam. From the geometrical properties of the box and the virtual structures we build a 3D computer model that is spatially registered to the framing marker set. The virtual structures were modeled in a way that the needle placement path was non-trivial with a limited choice of entry points that would actually lead to a successful needle placement.

As the needle is being tracked at all times, the distance of the needle to any of the virtual structures is calculated and monitored continuously. To gauge user performance we computed the following parameters:

- Distance between needle tip and center of lesion once the user believes he hit the target,
- Number of contacts of the needle with critical structures throughout the procedure,
- Duration of the procedure and the steps in the workflow,
- User's opinion of whether he believes he has been successful or not.

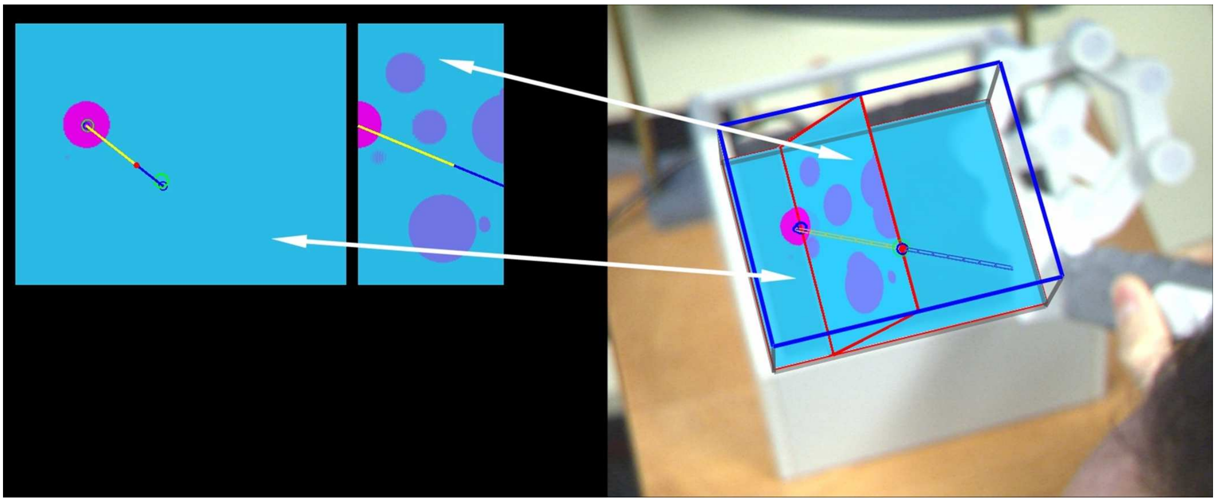


Figure 4.10: Visualization for 2D navigation (*left*) and 3D navigation (*right*). Two cross-sectional planes through the phantom box are displayed: one intersecting the target lesion and displaying the projection of the needle and its extension onto this plane, the other aligned with the needle and its extension. Both planes also show the cross-sections through simulated critical structures in the box. The 3D view additionally provides a live video view in the background of the graphics.

4.4.2 Four Visualization Schemes

We analyze four types of navigation methods, distinguished by the method of presenting the visual information to the user. The results of this study will show the extent of the influence of the visualization scheme on the outcome of the needle placement procedure. The four different visualization schemes are:

1. 2D visualization on a monitor screen,
2. Monoscopic 3D visualization on a monitor screen,
3. Stereoscopic visualization on a monitor screen with shutter glasses,
4. Stereoscopic visualization with a head-mounted display.

The first method is a 2D visualization method, whereas the latter three are 3D visualization methods. Methods 1 and 2 don't provide depth perception to the user. Method 3 provides a stereoscopic depth cue. Method 4 additionally provides a kinetic depth cue, which comes with the dynamic viewpoint changes of the head-mounted display.

All four visualization schemes present the same graphical information: two cross-sections through the phantom box, one plane intersecting the target lesion in a frontal view and one plane that contains the needle and its trajectory, as well as the needle and its extension in form of wireframe cylinders. Additionally the live video view of the real phantom box and needle is displayed in the background of the graphics for all 3D methods. Figure 4.10 shows the 2D view on the left and the 3D view on the right side. Methods 3 and 4 show the 3D view stereoscopically.

4.4.3 User Study and Results

We performed a user study with 28 participants, each of them utilizing all four navigation methods. Every user had the chance to try each method once before he was asked to perform five needle placements with lesions in different locations inside the phantom box. To compare all four navigation methods in regard of their visualization schemes, a specific workflow, which was the same for all methods, was given to the participants of the study:

1. Defining the target point by directing the graphical extension of the needle into the center of the simulated lesion on the cross-sectional cut plane.
2. Choosing an entry point on the surface of the foam by avoiding critical structures on the linear trajectory between entry point and previously defined target point.
3. Alignment of the virtual needle extension with the selected trajectory.
4. Insertion of the needle along the chosen path while avoiding critical structures.
5. Removal of the needle reversely along the same path while avoiding critical structures.

We consider the needle placement successful if at the end of the insertion the distance of the needle tip to the target center is less than 4mm and if no critical structures were touched throughout the procedure. Due to the optimized needle placement protocol, which changes the 3D task into a succession of 2D tasks, the study showed that all four navigation schemes results in almost the same result with respect to reaching the targets. The overall success was in between 80%–85% for all schemes. The four visualization methods showed differences in the study outcome with respect to avoiding critical structures, though. Both, the 2D approach (method 1) and the HMD approach (method 4), had a $>$80% success rate, whereas the monoscopic 3D approach (method 2), as well as the stereoscopic shutter glass approach (method 3), yielded a success rate below 70%.

The 2D visualization scheme provides direct views on both cross-sections through the phantom box and the HMD-based approach provides the user with the ability to move his head around the phantom and inspect the structure from different viewpoints to find an optimal needle path,

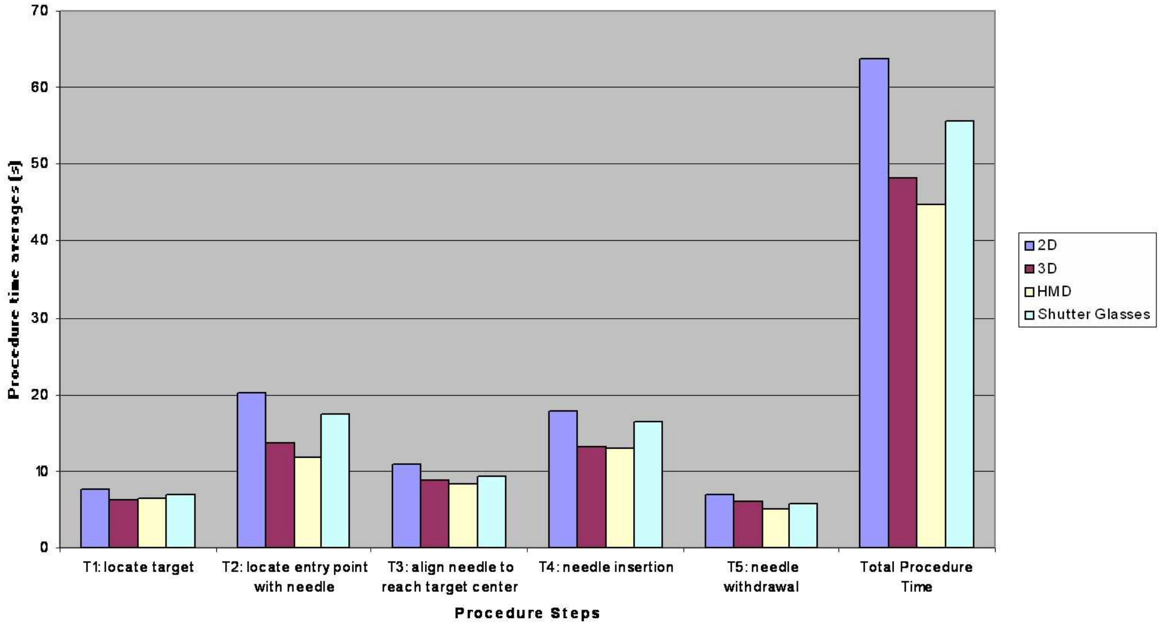


Figure 4.11: Time for performing the needle placement procedure with four different visualization schemes for navigational guidance.

a freedom which the shutter-glass visualization scheme lacks. The monoscopic 3D approach turned out to be more confusing to the user due to the missing depth perception and the uncertain spatial relationship between target and needle entry planes.

Procedure time is an important criteria in medical interventions, not only to minimize interventional complications and infections but also costs. Figure 4.11 displays the average time that was needed to perform the needle placement task for all four methods. The 3D guidance schemes (methods 2, 3, and 4) are performed faster than the 2D approach (method 1), which can mainly be attributed to the necessity of building a mental 3D model during a 2D guided task. The HMD method was the fastest (45s) and the 2D method the slowest (65s) in average.

In a medical scenario it is also important that the user can trust in his judgments based on the navigational guidance. As such, an important outcome of the study was the judgment of the user, whether he hit any critical structures or not, compared to the actual recordings of the needle in relation to the critical structures. The HMD method proved to be the most reliable, with the highest number of confirmations of the user's judgment. The shutter glass method was the least reliable, with a high number of overestimations, where the user believed in success, but the system actually recorded a failure.

Chapter 5

RAMP—Medical Applications and Pre-Clinical Evaluation

The capability of blending virtual structures in real-time seemingly realistically into the view of the user's environment without any jitter or swimming of the graphics creates a striking visual experience. The RAMP system provides a stable, tested, and technically evaluated foundation for bringing augmented reality a step closer to the area of surgical and minimally invasive interventional navigation. We propose the term *in-situ* visualization: anatomical structures are being displayed at the location where they actually are. The physician can see beyond the surface, the patient's body becomes transparent for him. This chapter introduces the RAMP system into different medical scenarios in various hospitals and evaluates its potential for providing valuable navigational support to the surgeon, interventionalist, radiologist, or physician.

5.1 Surgery Planning

We adapted the RAMP system to be used as a surgery planning or training tool, providing interactive in-situ visualization of 3D medical imaging data [Vog04a]. For high-quality rendering of the medical imaging information and the overall augmented scene we utilize the capabilities of the powerful graphics card of the system. Fast high-precision MPR generation and volume rendering is realized with OpenGL 3D textures. We provide a tracked hand-held tool to interact with the medical imaging data in its actual location. This tool is represented as a virtual tool in the space of the medical imaging data. The user can assign different functionality to it: select arbitrary MPR cross-sections, guide a local volume rendered cube through the medical data, change the transfer function, among others. The user's spatial perception is based on ste-

Figure 5.1: Setup for demonstration purposes of augmented reality visualization for surgery planning. The head phantom contains a real human skull and was CT scanned to provide internal anatomical information.

reoscopic depth cues as well as on the kinetic depth cues that he receives with the viewpoint variations and the interactive data visualization.—The physician can interactively explore the medical imaging information in-situ.

5.1.1 Setup

To demonstrate the principle of using augmented reality as a planing and training tool we utilized a head phantom that contains a real human skull. At a clinical site we scanned that head with a CT scanner and loaded all DICOM images into a dedicated RAMP application. Figure 5.1 shows the head on a platform that we built for this application. We constructed a framing set of retroreflective markers around the head phantom. If this setup were to be used on a real patient, we would need to attach the marker frame rigidly to the patient's head.

We registered the CT dataset to the coordinate system of the framing marker set by creating several 3D point correspondences between obvious anatomical surface structures in the CT volume dataset and the outer surface of the phantom. To measure points in 3D space we utilized a tracked pointer with the ART-tracking system in our lab that also tracked the pose of the marker frame. Once we found a rigid body transformation $\boldsymbol{H}_{\mathrm{fd}}$ (see Chapter 3.2.5) between the marker frame and the CT dataset this way, we could simply rely on our head-mounted tracker camera to provide correct spatial information of the CT dataset relative to the marker frame, and as such, relative to the head, which was put into a fixed position on the platform. We found the registration error between the virtual and real objects to be less than 1.5 mm.

5.1.2 In-Situ Visualization of Volumetric Medical Imaging Data

Visualizing volumetric datasets, such as 3D images from a CT scan, in real-time with 30 frames per second and stereoscopic XGA resolution is a challenging task. We implement the rendering of the augmented scenes with OpenGL, what we display is the view of the two OpenGL cameras. We have about 15ms to render one full augmented reality image in XGA resolution (1024×768). From that, the graphics card needs 3–5 ms for uploading and rendering the background video image.

For the visualization of the medical data set we are taking advantage of the OpenGL 3D texture capabilities of this graphics board. Once the whole 3D data set (a 256^3 cube) is transfered to the texture memory, one can specify arbitrary cross-sections and can map these on corresponding polygons, whose 3D coordinates are specified in the medical data coordinate system. This technique allows us to generate an MPR cross-section (see Chapter 2.4.2) with trilinear filtering and to render it into the real scene in less than 1 ms. We are furthermore taking advantage of the dependent texture lookup extension to generate the MPRs with post-classification (see [Eng02] and Chapter 2.4.2), which means that the transfer function is applied only *after* the cross-section is generated (with trilinear filtering) from the dataset.

We also integrated volume rendering into the RAMP system by rendering view-aligned MPR cross-sections (Chapter 2.4.2) from the 3D texture, resembling the CT dataset. If we sacrifice the 30Hz real-time performance then our AR system can augment the real scene by volume rendering of large data sets as shown in Figure 5.2 (with about 10 frames/second for stereo XGA). A stereoscopic, dynamic AR view of this in-situ volume rendered head phantom is quite impressive and will be possible in 30Hz real-time with newer generation graphics cards. Currently we have to limit volume rendering to a 64^3 crop-box. But, as the next section will show, the user is able to interactively move this crop-box through the whole data set, i.e., through the augmented

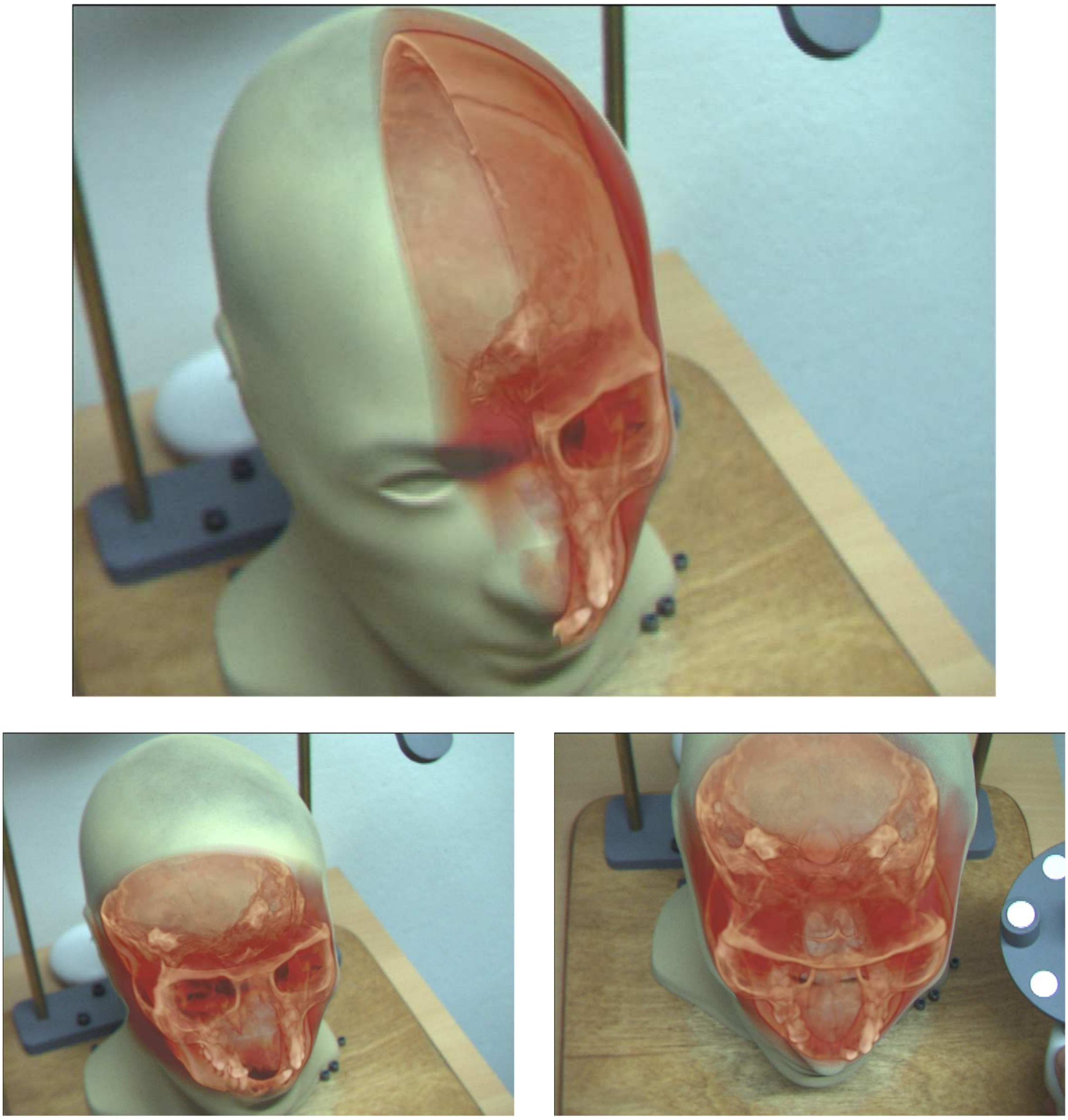


Figure 5.2: In-situ volume-rendering of a CT dataset inside the head phantom. The images show snapshots of the augmented video images that were created with the RAMP system. The stereoscopic and kinetic depth cues provided by the HMD cannot be captured in this figure. Through the HMD it seems as if the head phantom in front of the user were transparent.

phantom in front of him. Noteworthy is also that the post-classification technique with dependent texture lookups ensures high quality volume rendering. It furthermore makes it easy to change the transfer-function “on-the-fly.” In fact, the user can interactively change it while viewing.

Figure 5.3: *Left:* Wireless trackball with attached marker cluster. The 6D pose of this trackball is tracked by the head-mounted infrared camera. *Right:* Interactive in-situ visualization of anatomic structures inside a phantom head. ("Hands-on experience" at the Matrix exhibition at the European Congress of Radiology)

5.1.3 Interactivity with a Hand-Held Input Device

To add interactivity to the AR system, we engineered a hand-held tool whose pose can be tracked with the head-mounted tracker camera (as explained in Chapter 3.2.7). Figure 5.3 shows this modified wireless trackball, which is equipped with a set of retroreflective markers. Its tracked pose is used to select regions of interest and to browse through the medical dataset during the operation of the AR system. The user can freely move this tool along the side of the real object (or the patient). A virtual plane, which is "rigidly" attached to the tool, "cuts" through the medical dataset and generates the corresponding MPR cross-section. The user can freeze arbitrary MPR slices in space by using the push-buttons. Furthermore, the trackball itself and the buttons are used to provide a larger freedom in rotational operations and to change visualization settings, respectively. Pre-segmented structures from the medical data set can be integrated into the augmented view in form of wireframe or shaded models. Figure 5.3 shows on the right side how the CT scan of a head phantom can be explored in a very intuitive and direct way—inside the real head, as if the head were transparent.

Another visualization option is to guide a virtual pointer through the real object to apply volume rendering to parts of the 3D medical dataset. The virtual pointer is spatially attached to

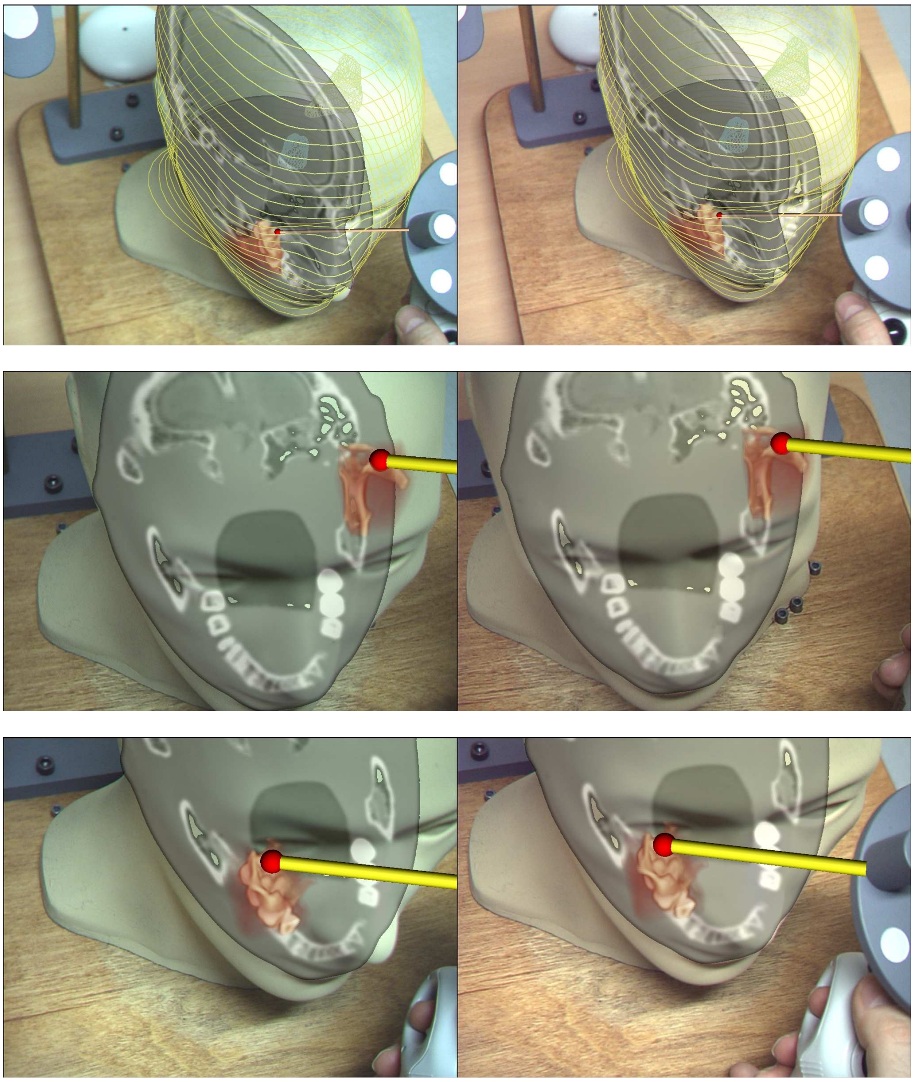


Figure 5.4: AR visualization of MPR cross-sections and interactive volume-rendering of a CT scan inside a phantom head. A virtual pointer is "rigidly" connected to an optically tracked hand-held input device which moves a volume rendered cube through the inside of the head. *(A "parallel stereo look" at these pictures helps to gain a better understanding of the necessity of the stereoscopic depth cue.)*

the hand-held tool and has a crop-box (of size 32^3) at its tip, which applies volume-rendering to this particular cubic section of the CT dataset. The user can change the transfer function and the transparency of the dataset with the trackball to emphasize certain tissue and bone structures. Figure 5.4 and Figure 5.5 show examples of the stereoscopic augmented view of this interactive visualization technique.

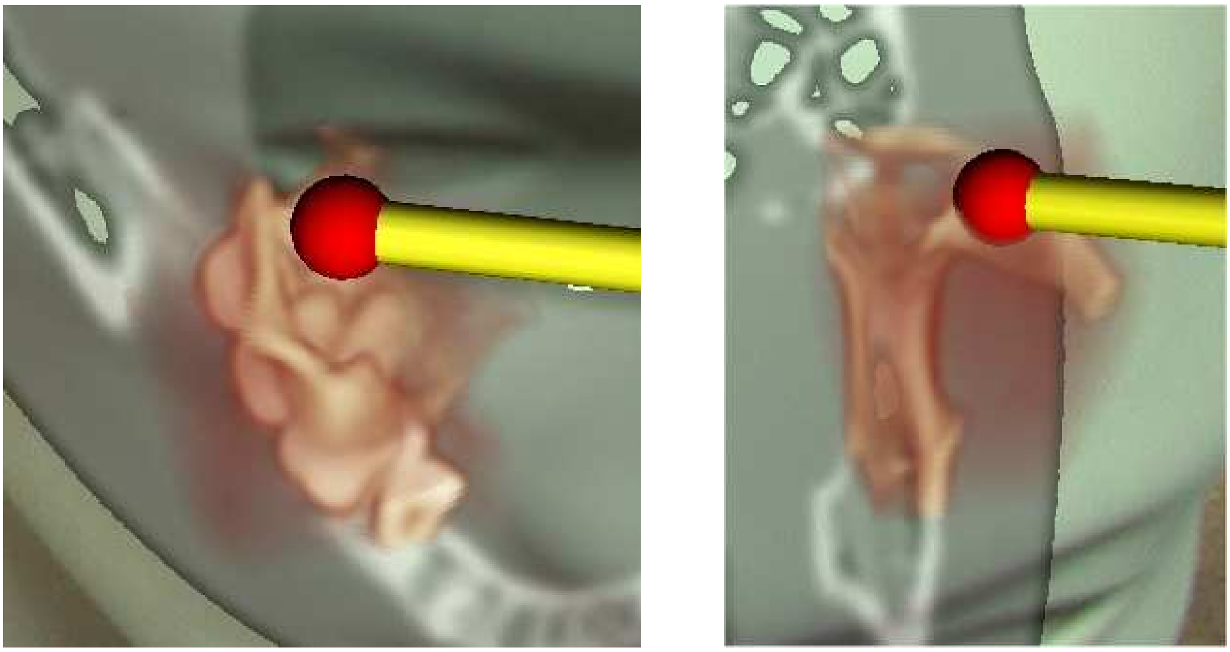


Figure 5.5: Zoomed view of in-situ volume rendered anatomical structures during interactive AR visualization. An underlying MPR cross-section guides the user to move the virtual pointer through the phantom head. A volume-rendered section of the head is always visible at the tip of the virtual pointer.

5.1.4 User Experience

The introduced application of the RAMP system for surgical/interventional planning and training purposes has been tested and evaluated by multiple radiologists and physicians. During the days of the European Congress of Radiology in Vienna in 2002 and 2003 it served as a "hands-on" exhibit and was highly frequented and referred among participants of the congress. The common feedback was that augmented reality visualization is able to provide a user interface to medical imaging information that is compellingly direct and intuitive and as such can help to plan and possibly perform interventional procedures more efficiently and precisely. The spatial relationship between the surface of a patient's head and its anatomy, as provided by medical images, can easily be assessed. For instance, the best site for opening the skull for a tumor resection can be planned by assessing all surrounding structures at once. The interactivity that the hand-held input device provides enables the user to explore 3D medical datasets in a direct way in the context of the actual patient. Features that made our system particularly suited for medical applications are the high (XGA) resolution of the display, the elimination of time lag between video and graphics, and the low latency of the visual feedback of only about 0.1 seconds.

Figure 5.6: In-situ visualization of MRI images on a phantom (melon) in a neurosurgical suite. *Left:* Sterilizable marker frame attached to the head clamp of an interventional MRI scanner bed (Siemens OpenViva). *Right:* Augmented view of the melon with two MRI cross-sections displayed in-situ.

5.2 MRI-Guided Neurosurgery

Neurosurgery is an area where image guidance systems are already employed on a regular basis. A mapping between the medical images and the patient is displayed on a separate screen, usually by means of spatially tracked instruments and pointers. As such, the positions of the instruments are presented in the context of the patient's head scans, separate from the actual physical body. The resection of a brain tumor, for instance, requires a precise mapping of those images onto the patient's head and opened skull—a mental mapping that experienced neurosurgeons have to perform during the procedure. Augmented reality provides the opportunity to visualize medical imaging information and surgical tools within the patient—*in-situ*, a concept that sounds especially promising for neurosurgical navigation. In [Sau02b, Wen03, Sau01a] we adapted the RAMP system to an MRI neurosurgical suite at UCLA (University of California, Los Angeles) and report on pre-clinical evaluations on phantoms. Later on, we visually evaluated the system with a neurosurgeon preceding an actual neurosurgical procedure at HUP (Hospital of the University of Pennsylvania, Philadelphia).

5.2.1 AR for a Neurosurgical MRI Operating Suite

To be able to study the RAMP system in a neurosurgical setup, we adapted the system to the interventional MRI operating room at neurosurgical department at UCLA. During the time of our studies it contained a Siemens OpenViva 0.2 Tesla MRI scanner. The surgical table provides

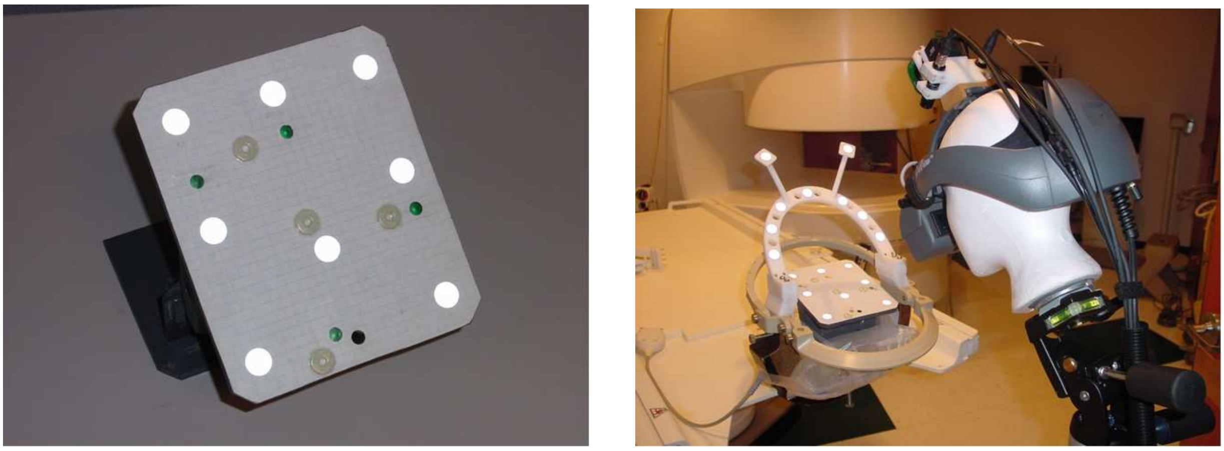


Figure 5.7: *Left:* Calibration phantom with retroreflective markers and MRI markers. *Right:* Calibration phantom within the neurosurgical head clamp of an interventional MRI suite.

a mechanism that can bring the table with the patient into the MRI scanner for scanning and then swivel him out into the fringe field of the magnet to continue with the surgical procedure. We established a direct connection between the RAMP system and the scanners console to transfer the medical DICOM images directly into our software. We also implemented a simple software tool to interactively extract 3D wireframe models—representing tumors and other structures—from the MRI data.

During a neurosurgical procedure the patient's head will usually be stabilized and rigidly attached (by screws that penetrate into the top surface of the skull) to a clamp at the end of the surgical table. We built a sterilizable frame of retroreflective markers that can easily be installed on top of this head clamp. Figure 5.6 shows this setup and the MRI scanner in the background. The marker frame can be taken on and off the head clamp and always snaps reproducibly into its correct position. Since the MRI scans are taken during the intervention, while the patient is affixed to the head clamp, the patient's head does not move with respect to the marker frame either. This simplifies the registration process between the coordinate systems of the medical data and the marker frame, i.e., determining the rigid body transformation $\boldsymbol{H}_{\mathrm{fd}}$ (see Chapter 3.2.5).

The scanner precisely keeps track of the movements of the scanner table (as most CT and MRI scanners do) and rigidly "attaches" its scanner coordinate system to the table. Therefore, every MRI image taken throughout the intervention has known coordinates with respect to the head clamp. To find the rigid body transformation between the marker frame coordinate system and the scanner coordinate system—once, before the intervention takes place—we built a calibration phantom that contains retroreflective markers and MRI markers (Figure 5.7). This initial calibration is performed in a setup depicted on the right side of Figure 5.7, with the

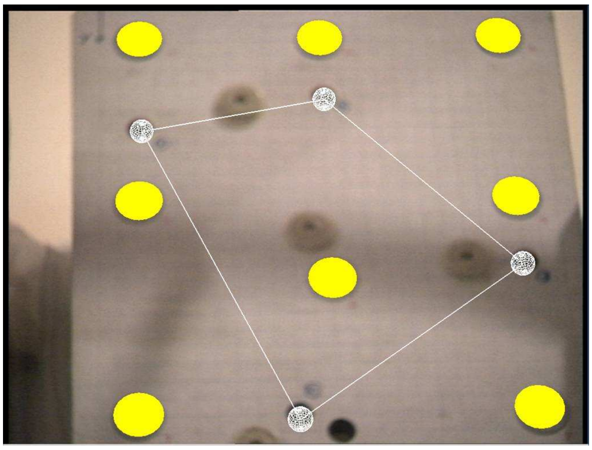


Figure 5.8: AR control view of the calibration phantom after calibration. The retroreflective markers are overlaid with virtual discs and the MRI markers with virtual spheres of corresponding sizes.

calibration phantom placed inside the surgical head clamp. To estimate the 3D position of the retroreflective markers of the calibration phantom with respect to the marker frame, the head-mounted tracker camera captures a single image of the marker frame and the calibration phantom. This automatically gives us the MRI marker coordinates of the phantom in the marker frame coordinate system as well. Then, we take the marker frame off the head clamp, swivel the scanner table into the MRI scanner, and scan the phantom. The scanner coordinates of the spherical MRI markers can be extracted from the MRI dataset. Since now the coordinates of the MRI markers are known in the scanner coordinate system and marker frame coordinate system, a simple optimization routine yields the rigid body transformation $\boldsymbol{H}_{\mathrm{fd}}$ between both coordinate systems. This transformation stays constant throughout the intervention and can be applied to map the interventional MRI data into the patient's head. Figure 5.8 shows a control view of the calibration phantom that is augmented with graphical representations of the retroreflective and MRI markers.

5.2.2 Phantom Experiments

We performed three types of experiments in the AR-adapted interventional MRI operating suite. For the first experiment we implanted several vitamin E capsules into a cantaloupe. The neurosurgeon's task was to remove those capsules with a rongeur in a minimally invasive way. We utilized an MRI scan of the cantaloupe, after it was placed into the neurosurgical head clamp, to generate 3D wireframe surface models of the vitamin E capsules inside the cantaloupe. Figure 5.9 shows

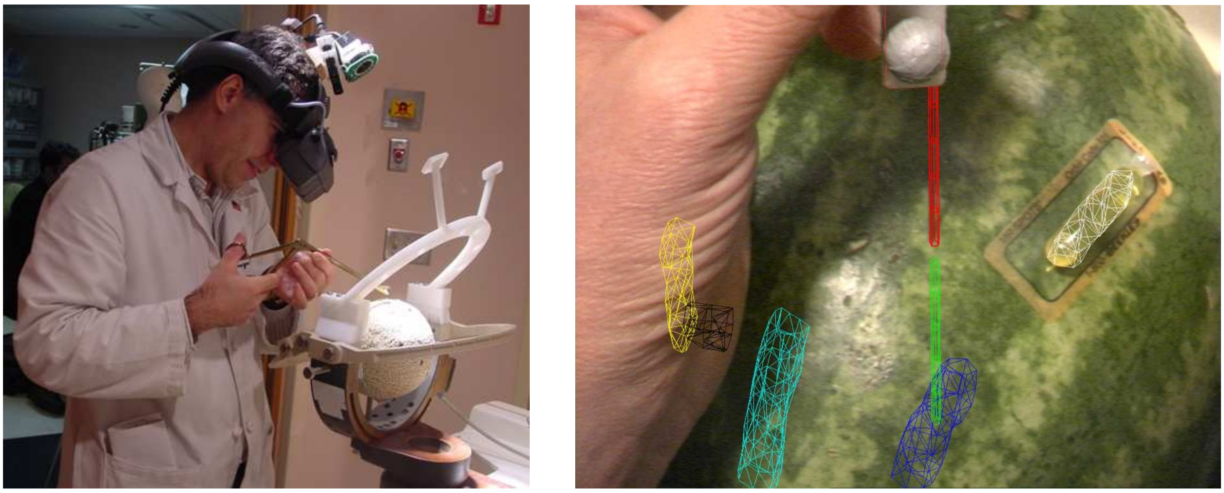


Figure 5.9: Phantom experiments in the interventional MRI suite at UCLA. *Left:* Dr. Rubino is extracting implanted vitamin E capsules from a cantaloupe with a rongeur under AR guidance. *Right:* AR view of a tracked needle with extrapolated needle path which is used to target vitamin E capsules (represented by wireframe models) inside a water melon.

Dr. Rubino on the left, performing this AR guided experiment. The augmented view, with its stereoscopic and kinetic depth cues, of the vitamin E capsules provided good guidance to extract the capsules that were close to the surface of the phantom (within 1–3cm underneath the surface). The perceived spatial relationship between the wireframe models and the outer shell of the phantom enabled the surgeon to chose an appropriate entry point for the rongeur and to extract the capsule minimally invasively. Since we don't track the rongeur itself, the instrument becomes invisible to the surgeon's eye, from the point it enters the phantom. Therefore, the deeper lying targets (8–10 cm underneath the surface of the cantaloupe) were hard to extract this way.

Since it was impossible to generate a graphical model of the rongeur itself during the experiments, the second experiment involved a tracked needle that the surgeon placed into vitamin E targets inside a water melon. Figure 5.9 shows an augmented view of this setup on the right side. The targets are represented as wireframe models and the needle as a thin wireframe cylinder. The extrapolated needle path, which was "rigidly" attached to the needle, guided the surgeon to chose an appropriate entry point for the deep lying targets. Therefore, the user would see where the extrapolated needle path would intersect with the target. Furthermore, all targets could be reached easily since the graphical representation of the needle and the targets made the real melon itself look transparent. All targets (7mm wide) were hit consistently hit as deep as 5cm inside the phantom. Due to the length of the needle itself the experiment was restricted to 5cm deep targets.

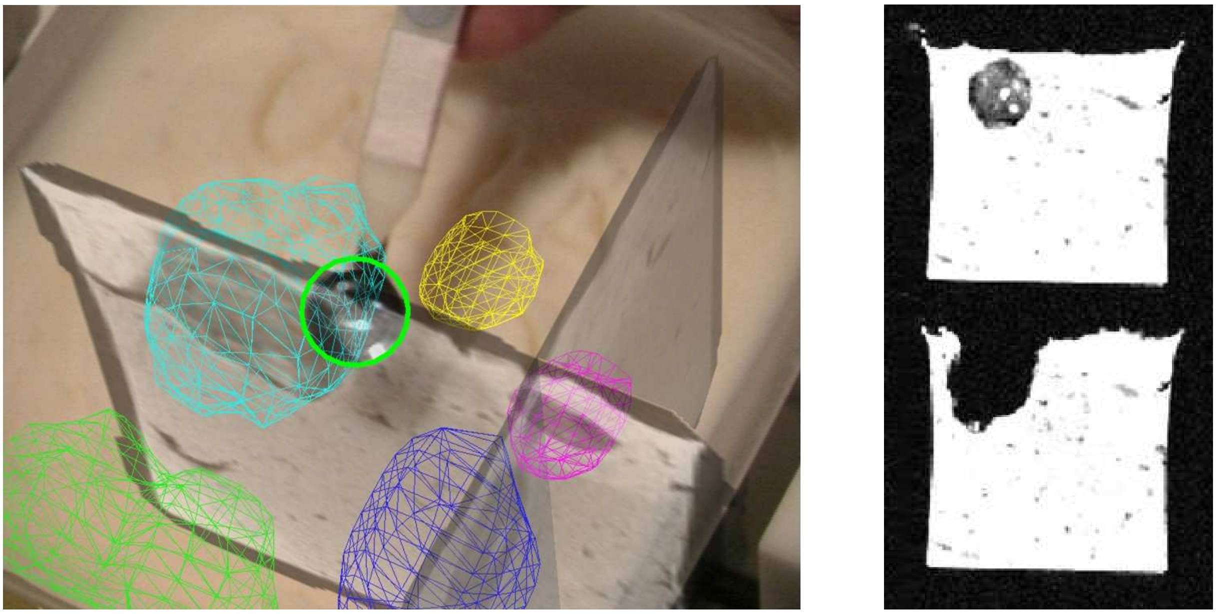


Figure 5.10: Experimental AR-guided “tumor resection” with a gadolinium-doped mashed potato phantom. *Left:* Augmented view of the phantom with two orthogonal MRI cross-sectional slices and wireframe models for the gadolinium-doped potato parts (tumor targets). The instrument—a tracked spoon—is graphically represented by a circle and visualizes the interaction of the spoon with the 3D models of the targets. (However, this image alone does not provide the necessary stereoscopic and kinetic cues to perceive depth correctly, but the RAMP system does.) *Right:* MRI scan of the phantom before *(top)* and after *(bottom)* the resection of one of the targets.

To simulate a tumor resection we designed a third experiment utilizing a phantom box containing mashed potatoes. The tumor was simulated by mixing the MRI-contrast agent gadolinium into some confined parts of the mashed potatoes inside the phantom box. As such, direct inspection would not be sufficient to distinguish the different types of “tissue,” but an MRI scan would be. We equipped a measuring spoon with a marker cluster to be able to track this instrument for the task of resecting (scooping out) the gadolinium doped mashed potato parts under AR guidance. Figure 5.10 shows an augmented view taken during this experiment. Due to the depth cues provided by the RAMP system the surgeon is able to locate the selected MRI slices and the 3D wireframe models of the targets inside the actual phantom. Also, the spoon itself is represented as a circle and helps once the actual instrument disappears underneath the surface of the phantom.

The RAMP system provided AR guidance while the neurosurgeon performed four tumor resections with the phantom setup. The right side of Figure 5.10 shows images of MRI scans

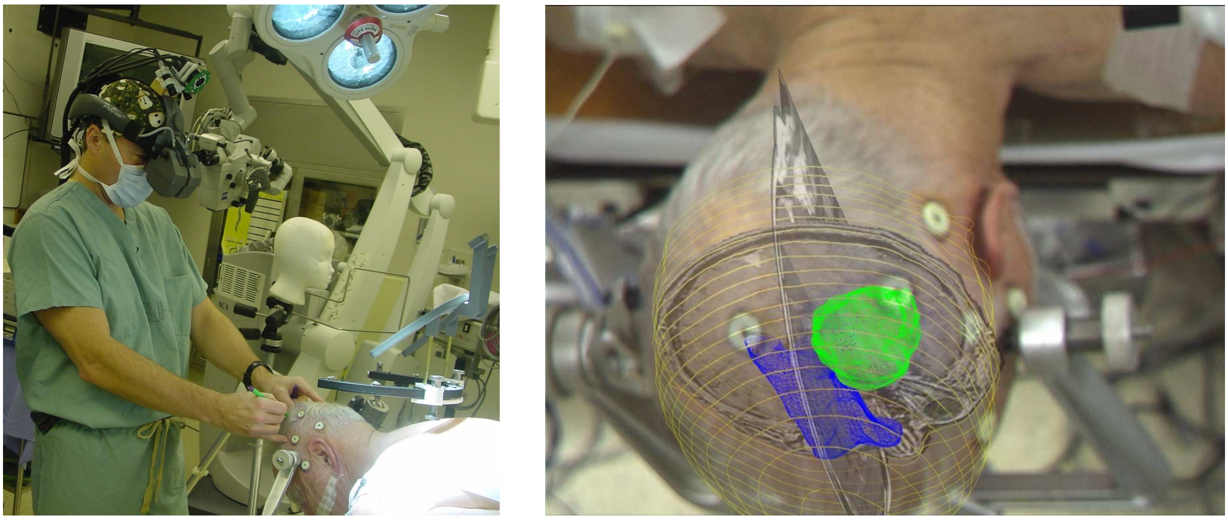


Figure 5.11: The RAMP system is used at the Hospital of the University of Pennsylvania to collect feedback of the three-dimensional mapping of MRI data and the actual patient before the craniotomy. *Left:* Dr. Grady inspecting the patient under AR guidance. *Right:* Patient's head augmented with MRI slices and wireframe models of critical tissue in the brain.

taken before (pre-operatively) and after (post-operatively) the resection task. Neglecting the top layer of the phantom in our calculations (since the potato part above the target obviously could not remain in the phantom during this task), on average 40% more material was extracted than would have been neccessary for the removal of the complete target. Priority to the neurosurgeon was to extract all possible remains of contrast enhanced material. Moreover, due to the consistency of the phantom (sticky mashed potatoes), it was impossible to precisely extract along the borderline of the targets. Even though the phantom and the instruments were sub-optimal, the experiment demonstrated the capabilities of the AR system in guiding neurosurgical procedures intuitively and effectively.

5.2.3 Guidance Preceding a Craniotomy

Besides UCLA, we were able to install the RAMP system at the Hospital of Pennsylvania, preceding an actual neurosurgical intervention. This setup was solely used to collect feedback from the neurosurgeon before the actual intervention. (The RAMP prototype system is not an approved device for clinical use.) Figure 5.11 and Figure 5.12 show the patient inside the neurosurgical head clamp of the operating table. Dr. Grady, the neurosurgeon, is using the RAMP system to give us feedback about his planed craniotomy (opening of the skull).

In this setup the MRI scanner was not located inside the operating room, but the patient was

Figure 5.12: First impressions of the potential of AR guidance preceding a neurosurgical intervention at the Hospital of the University of Pennsylvania. Unlike the UCLA configuration, this operating room does not contain the MRI scanner, but a pre-operative MRI scan in conjunction with skin markers is used to augment the patient's head.

scanned beforehand, with MRI skin markers on his skull. Once the patient was placed inside the neurosurgical head clamp we measured those markers with a tracked pointer inside the coordinate system of the marker frame on top of the clamp. Since those markers could be extracted from the MRI scan, we could find a transformation $\boldsymbol{H}_{\mathrm{fd}}$ relating data and marker frame coordinate systems with each other. (Typically, once the patient's skull gets opened, a brain shift occurs, which invalidates the registration of the pre-operative MRI scan.)

Using the RAMP system, the neurosurgeon gained a good understanding of the spatial relationship of the patient's anatomy and his head. In fact, the craniotomy could have been planned intuitively and effectively utilizing the RAMP system. The essence of the received feedback was that *in-situ* visualization enables an unprecedented intuitive three-dimensional mapping between the patient's MRI data and his head and could provide effective guidance for many types of neurosurgical procedures.

5.3 CT- and MRI-Guidance in Interventional Radiology

Interventionally acquired imaging information is an essential requirement for minimally invasive procedures. Fluoroscopy is capable of providing real-time updates, but only as a 2D projection image and with the cost of radiation to the interventional radiologist and the patient throughout the procedure (see Chapter 2.3.1). MRI and CT imaging information is utilized to guide minimally invasive interventions since it provides high-resolution 3D images and the ability to distinguish different tissue types (see Chapter 2.3.2 and Chapter 2.3.3). MRI scanners with closed magnets are more widely available and deliver a higher image quality than MRI scanners with open magnets. CT scanners are available in many radiology departments. When using CT scanners or MRI scanners with closed magnets, the procedure is usually performed in an iterative way due to the confined space inside the scanner. The patient is moved out of the scanner to advance the interventional instrument and moved into the scanner to acquire new images. This *stop-and-go* method ensures stepwise monitoring of the procedure, but requires many interruptions of the actual intervention. Examples for these types of interventions are ablations of liver tumors and abdominal needle biopsies.

CT fluoroscopy guidance inside the CT scanner gantry avoids this stop-and-go method but for a price of extended radiation exposure to the patient and the operator. MRI scanners with closed magnets usually don't provide enough space to perform an intervention inside their long bore hole. Augmented reality has the potential to intuitively guide interventions directly on the scanner table, without the need for a stop-and-go technique or extended fluoroscopic radiation exposure. A technical evaluation of AR-guided needle placement tasks was given in Chapter 4. This section describes phantom and animal studies that we performed with the RAMP system for needle placement procedures under CT guidance at Brigham and Women's Hospital in Boston, Massachusetts [Sau03, Das06, Sch02] and under MRI guidance at the University Hospitals of Cleveland, Ohio [Vog04b, Wac06, Kha03b].

5.3.1 AR for CT-Guided Needle Procedures

One of the potential applications of augmented reality is its use as an image-guided navigation technique for needle placement procedures alternatively or complimentary to CT fluoroscopy guidance, where the needle is monitored under x-ray [Sil99]. Compared to fluoroscopy, AR guidance utilizes recorded and registered CT images in combination with a tracked needle, instead of real-time x-ray imaging. As such, radiation exposure to both operator and patient can be reduced. Furthermore, CT fluoroscopy guidance requires in-plane needle placement for efficient

Figure 5.13: Interventional CT Suite at Brigham and Women's Hospital. *Left:* Suite equipped with the RAMP system on a separate cart and the marker frame on top of the scanner table. *Right:* Calibration phantom, containing retroreflective and CT markers, placed inside the marker frame setup.

monitoring, which restricts the needle path to be more or less orthogonal to the patient's head-toe axis. AR can provide guidance for off-plane and double-oblique access paths as well. Correct registration between the recorded CT images and the patient has to be provided to guarantee successful AR guidance, of course.

In Chapter 4.3 we preformed a technical evaluation of the RAMP system for needle placement procedures on a phantom box with hidden push buttons of 6mm diameter. The following describes an adaptation of the needle-guidance application with the RAMP system to an interventional CT suite at Brigham and Women's Hospital in Boston, Massachusetts. In a clinical setting we performed AR guided needle placement experiments on an interventional abdominal phantom that contained simulated liver lesions. The accuracy of each needle placement was determined with a control CT scan.

The interventional CT suite at Brigham and Women's Hospital—equipped with the RAMP system—is shown in Figure 5.13. Unlike the MRI scanner table of the neurosurgical suites, where we could install a marker frame atop the head-clamp as shown in Chapter 5.2, we utilize the marker frame setup as introduced in Chapter 5.1. We place its wooden base rigidly onto the scanner table to provide a limited workspace that is framed by the retroreflective markers.

To calibrate the system we need to determine the transformation $\boldsymbol{H}_{\mathrm{fd}}$ between the scanner and the marker frame coordinate systems. For this purpose we use the same technique as described in Chapter 5.2.1, employing a phantom with eight retroreflective and four CT markers (in place of the MRI markers) in a known geometry. We determine the pose of the calibration phantom in the

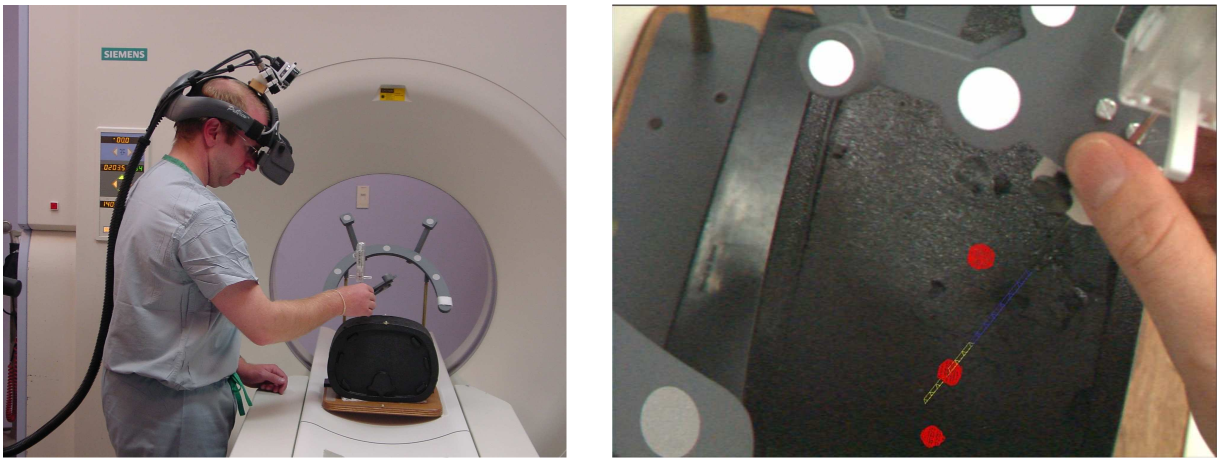


Figure 5.14: AR guided needle placement studies in an AR enabled CT suite at Brigham and Women's Hospital. *Left:* Dr. Schoepf targeting one of the simulated lesions inside an interventional abdominal phantom. *Right:* Augmented view of the abdominal phantom with graphical representations of the lesions, the inserted needle, and the virtual extension of the needle.

scanner coordinate system from a CT scan and its pose in the marker frame coordinate system from a single tracker camera image, which, combined, gives us the needed transformation.

To guide needle placement procedures, we attach a marker cluster to a biopsy needle, similar to previous needle experiments, and provide a wireframe cylinder in the augmented view, representing the needle and its virtual forward extension. Segmented CT images provide the basis for the AR image guided needle placement task. To enhance the targets in the augmented view, we don't show the CT scan itself, but 3D wireframe models of the segmented lesions.

5.3.2 Phantom Study under CT-Guidance

We utilize an interventional 3D abdominal phantom (CIRS, Norfolk, VA) for our studies, which we place underneath the marker frame on the scanner table. The phantom contains several simulated liver lesions, where we select the ones of about 1cm diameter. Ten needle insertions were performed altogether by three users in this clinical setup. Figure 5.14 shows a picture of Dr. Schoepf inserting the tracked needle and a monoscopic snapshot of the stereoscopic augmented view. Supported by the stereoscopic and kinetic depth cues, the operator sees the wireframe models of selected lesions *in-situ*, i.e., in their actual location. The virtual extension of the tracked needle guides the user in choosing an appropriate entry point at the phantom surface and in pushing the needle along its path forward until it hits the target. Since the actual needle disappears from the operators view as soon as it enters the phantom, the graphical rep-

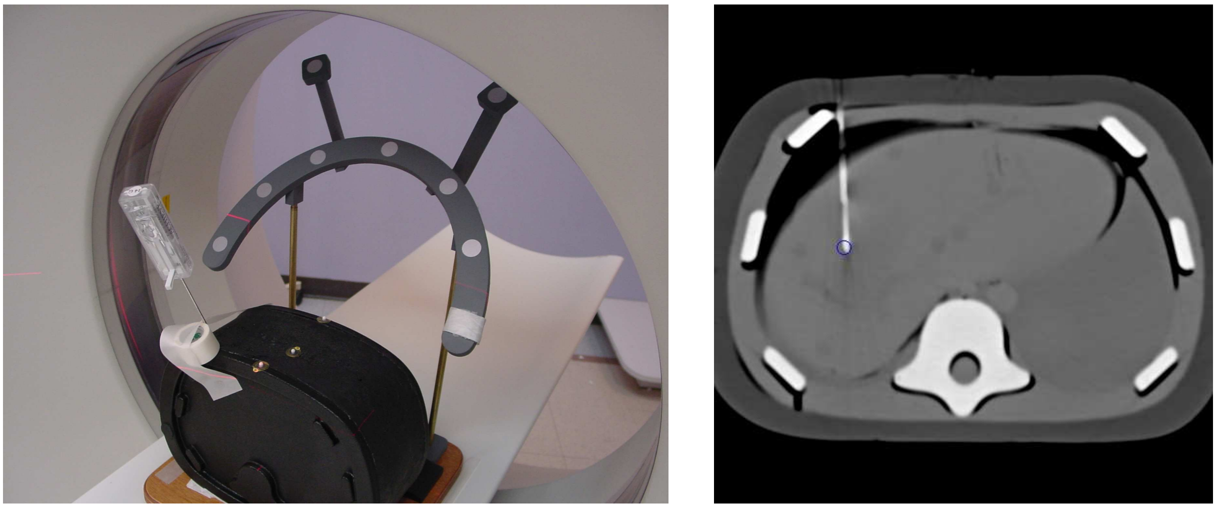


Figure 5.15: Performing a control CT scan of the phantom with inserted needle. *Left:* After insertion, the marker cluster is clipped off the needle and adhesive tape is used to stabilize the needle in its final position. *Right:* Axial CT slice of the phantom, showing the needle inserted into the lesion (marked with a ring).

resentation of the needle and its virtual extension are the ones that guide the user throughout the task. Aligning the wireframe cylinder of the needle extension with the wireframe target is technically all the user has to do to succeed while advancing the needle.

All users described the AR guidance as very intuitive and completed every needle insertion easily within 10 seconds. To evaluate the accuracy of the needle placement quantitatively we left the needle inserted after each procedure, detached the marker cluster from the needle, and stabilized it with some tape (left of Figure 5.15). Then we moved the scanner table back into the scanner to perform a control CT scan of the abdominal phantom. Figure 5.15 shows an axial CT slice through the phantom, with the needle being the bright line and ending inside the lesion, which we marked with a ring here due to its weak presence in the image. The CT scan confirmed that the needle was correctly placed into the lesion. We evaluated the needle placement experiment with respect to three different errors:

- *Virtual placement error (user error):* This error is computed from the tracking data and represents the distance between the virtual needle tip and the center of the lesion. It is a measure for how well the physician followed the AR guidance.

- *Real placement error:* This error is computed from the CT control scan and represents the distance between the real needle tip and the center of the lesion. It is a measure for the actual success of the needle placement procedure.

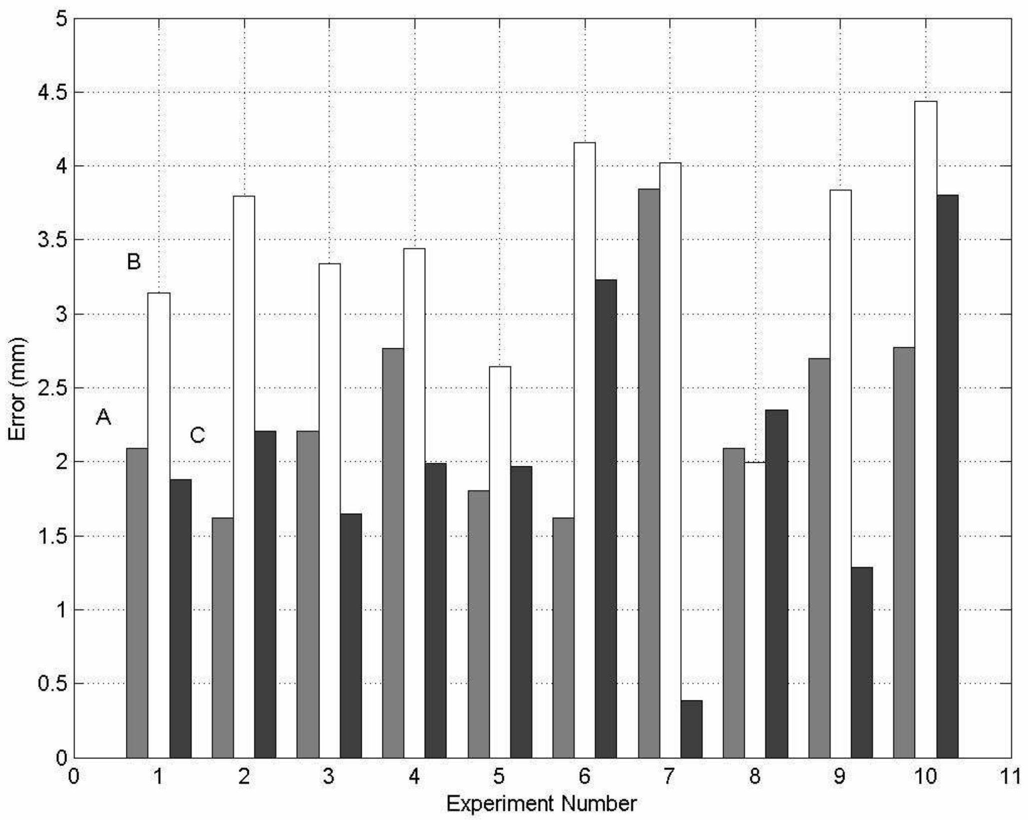


Figure 5.16: Needle placement errors of a set of 10 AR-guided needle placement experiments. *A:* Virtual placement error. *B:* Real placement error. *C:* AR system error.

- *AR system error:* This error is computed from the coordinates of the virtual and the real needle with respect to the lesion center and represents the distance between virtual and real needle tip. Assuming correct registration (and $\boldsymbol{H}_{\mathrm{fd}}$ calibration) between the virtual and the real lesion, this is a measure of how well the system tracks the pose of the needle.

The success of the actual needle placement, quantified by the real placement error, is dependent on both user performance and system performance. Hence, the real placement error is expected to be the biggest of all three errors. Figure 5.16 confirms this, which shows the errors for each individual experiment. The spatial distribution of the errors in respect to the target center for each needle placement is illustrated in Figure 5.17 and Figure 5.18. Finally, Figure 5.19 shows the averaged errors over all 10 experiments. The averaged real placement error is 3.5mm for all targeted liver lesions which have a diameter of about 10mm. The CT control scans confirmed that all ten targeted lesions were successfully hit.

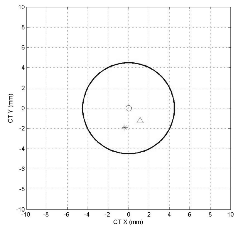
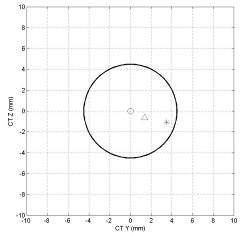
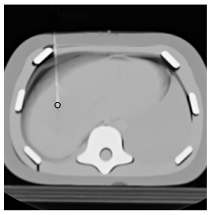
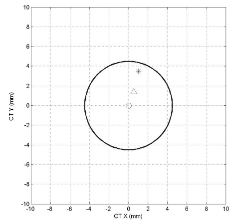
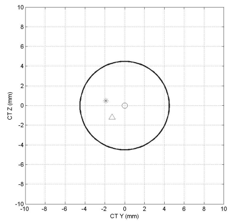
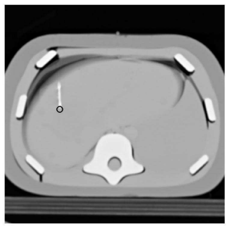
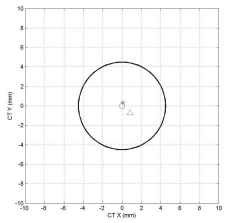
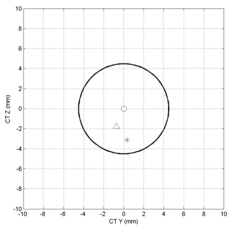
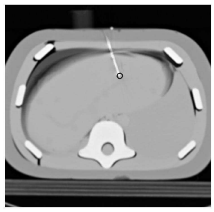
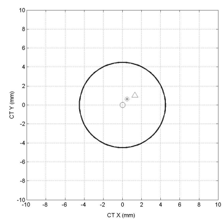
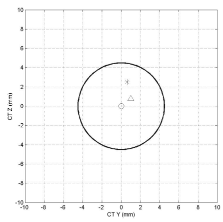
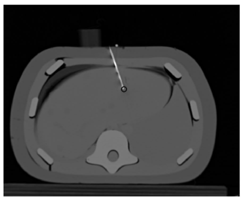
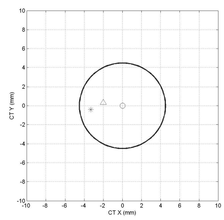
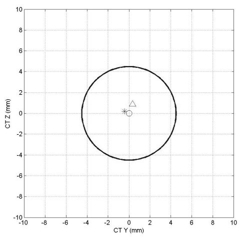
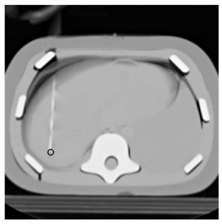
Experiment #	Axial View of the Target	Sagital View of the Target	Axial CT slice of the phantom (through the lesion center)
1			
2			
3			
4			
5			

Figure 5.17: Needle placement errors in respect to each target center (for experiments 1–5). The location of the real needle tip is represented by a star and the virtual needle tip by a triangle.

The average AR system error between virtual and real needle tip is about 2mm and comprises possible needle bending together with other calibration and registration errors. Since we used an 18 gauge needle, bending of the needle did not play a major role. Nevertheless, it is of concern

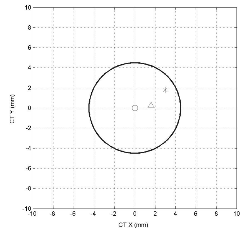
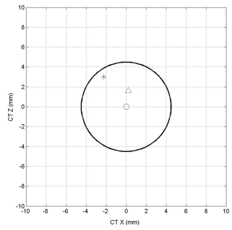
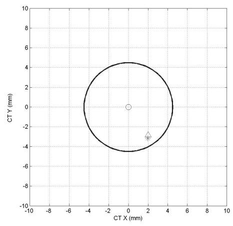
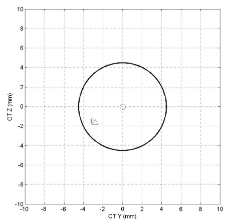
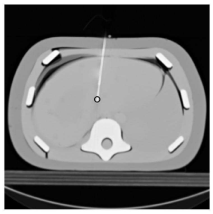
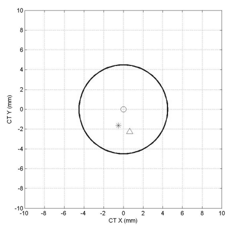
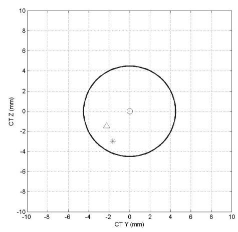
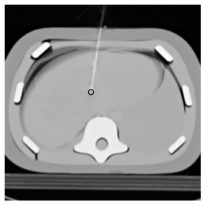
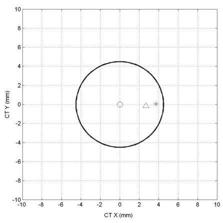
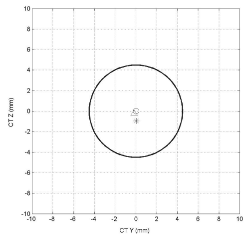
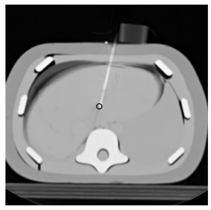
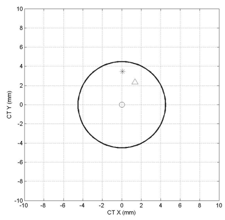
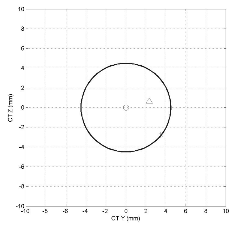
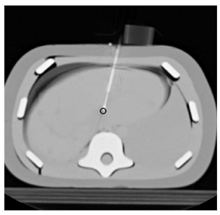
Experiment #	Axial View of the Target	Sagital View of the Target	Axial CT slice of the phantom (through the lesion center)
6			
7			
8			
9			
10			

Figure 5.18: Needle placement errors in respect to each target center (for experiments 6–10). The location of the real needle tip is represented by a star and the virtual needle tip by a triangle.

when using thinner needles, as the implemented tracking method is based on a marker cluster that is attached to the rear part of the needle. Since we assume a rigid virtual cylinder, attached to this marker cluster, that represents the needle and its forward extension, a needle that bends

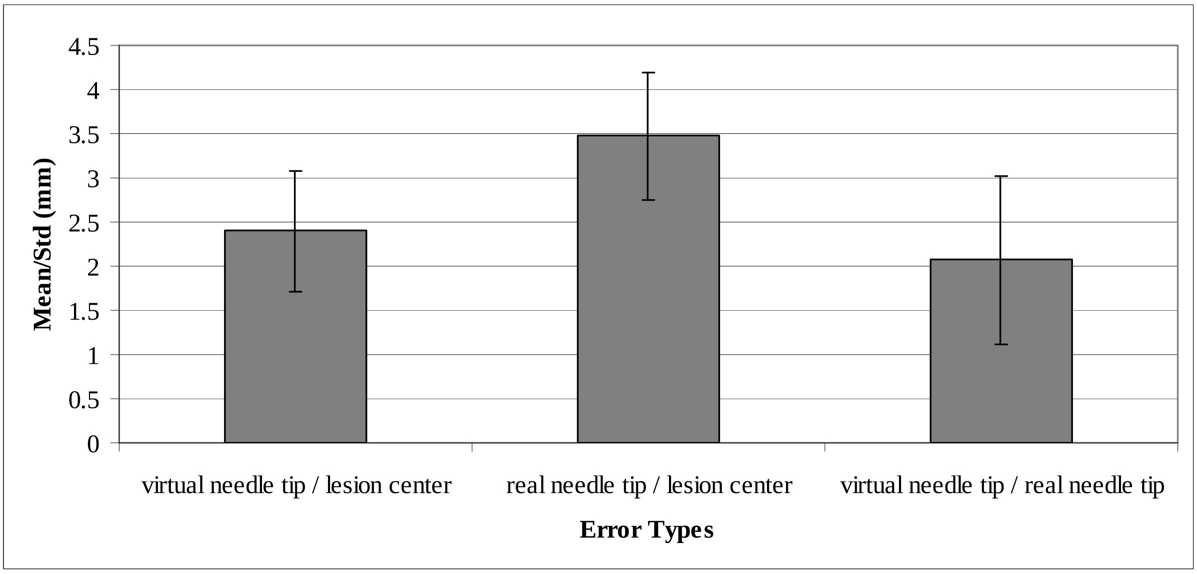


Figure 5.19: Needle placement errors averaged over 10 experiments. Both imprecise user performance *(left)* and imprecise system performance *(right)* contribute to the inaccuracy of the actual needle placement *(center)*.

during the insertion process would partially misguide the physician. As such, only a fairly rigid needle can be used with this tracking method to limit the detrimental influence of needle bending. Also the skills of using the guidance system and in performing needle insertions play a role. If the user choses an entry point carefully and aims into the correct direction before entering the needle into the surface of the phantom, the user can insert the needle along a very straight path and there is no need for bending the needle as means of correcting the path while approaching the lesion with the needle tip.

The user error, with an average of 2.5mm, seems to be quite high at first sight. In fact, one would assume that the user should be able to follow the visual guidance better than that. But during the current experiments the user was asked only to actually hit the lesion as good as possible, which itself was represented as a wireframe surface model. For that reason, the target center of this 10mm-sized lesion was not clearly pointed out and left to the judgment of the user. Future experiments should look into more defined representations of the target center to decrease the user error. For the tasks in this study, the 2.5mm virtual placement error is not bad, in fact. Certainly, the virtual needle tip has been successfully placed inside all wireframed lesions.

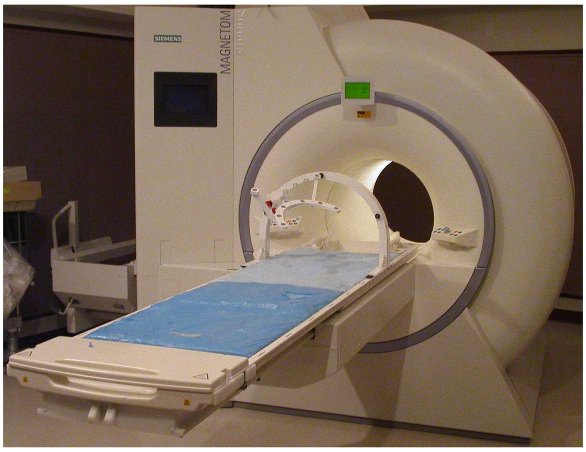
This study shows that AR guidance has the potential to decrease radiation dose to the patient and the physician for prolonged and complicated procedures where a CT fluoroscopic approach is unfavorable. Also, an intervention with complex anatomy can be guided more intuitively and is not restricted by the CT fluoro plane. Additionally, vulnerable vascular structures and perfused lesions might only be visible during distinct phases of a contrast enhanced CT scan—information that the AR system can easily integrate into the guiding graphics. Furthermore, since the AR guidance is given outside the CT scanner gantry there is no spatial constraint as with CT fluoroscopy guidance, which often impedes the use of longer biopsy needles or tumor ablation probes that are required to reach lesions deep inside the abdomen.

5.3.3 AR for MRI-Guided Needle Procedures

MRI imaging receives increasing attention for the guidance of interventional procedures. It does not emit harmful x-ray radiation to the patient or physician and, in general, provides better image contrast in soft tissue structures than CT. Open MRI scanners, such as C-shaped (Siemens) or double-doughnut shaped (GE) devices, allow the physician access to the patient inside the magnet during imaging. This provides immediate real-time MRI feedback during the insertion of biopsy needles or applicators for thermal ablation. However, these MRI scanners are not widely available and, additionally, only provide sub-optimal image quality due to the lower field strength compared to the more common closed-bore magnets. Those closed-bore MRI scanners can be found in many hospitals of the western world and provide excellent image quality based on a high field strength. The major challenge for interventional guidance is their long bore hole, though. To provide feedback for needle placement procedures the patient has to be moved back and forth between the position inside the bore where images can be acquired and a position outside the bore where the needle can be manipulated by the physician.

This time-consuming *stop-and-go* technique can be prevented by AR image guidance outside the bore hole, similar to Chapter 5.3.1. As with the CT approach, correct registration between the recorded MRI images and the patient has to be provided to guarantee successful AR guidance. The following describes an adaptation of the needle-guidance application with the RAMP system to an MRI suite at the University Hospitals of Cleveland at Case Western Reserve University in Cleveland, Ohio, containing a 1.5T closed-bore MRI scanner (Magnetom Sonata, Siemens Medical Solutions, Erlangen, Germany). The computer of the AR system was connected to the console computer of the MRI scanner by an ethernet connection to provide the MRI images for AR guidance. We performed AR guided needle placement experiments on phantoms and animals. The accuracy of each needle placement was determined with a control MRI scan.

Figure 5.20 shows the modified MRI scanner. In progress to our earlier experiments under CT guidance, we assembled a marker frame that attaches rigidly to the table of the MRI scanner, thereby establishing a reference for the patient coordinate system. The table, together with the



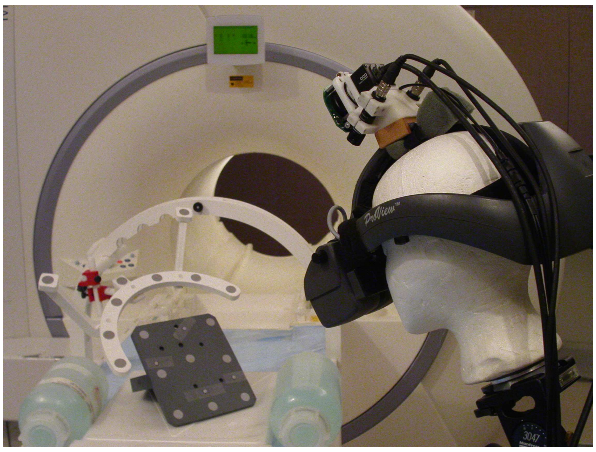


Figure 5.20: MRI suite at University Hospitals of Cleveland, Case Western Reserve University. *Left:* Marker frame which is rigidly attached to the table of a closed magnet Siemens MRI scanner. *Right:* Calibration phantom placed in workspace to perform a registration between the MRI scanner coordinate system and the marker frame coordinate system.

marker frame, can be moved in and out of the bore hole without changing the setup. The medical data coordinate system is given by the MRI scanner and remains fixed throughout the whole procedure. The target structures are extracted from MRI scans. For now, we assume that those structures don't move (with respect to the scanner table) after the scan was taken.

We utilize the calibration phantom from Chapter 5.2.1 and Chapter 5.3.1 to determine the rigid body transformation $\boldsymbol{H}_{\mathrm{fd}}$ that maps 3D points from the scanner coordinate system to the coordinate system of the marker frame. We place the calibration phantom, which contains eight retroreflective and four MRI markers in a known geometry, in the workspace (Figure 5.20 right). The pose of the calibration phantom with respect to the marker frame is determined with our tracker camera using the locations of the retroreflective markers, the pose of the calibration phantom in the MRI patient coordinate system is determined by finding the MRI markers in a scan of the phantom. A high-spatial-resolution MRI dataset was acquired for this purpose (three-dimensional fast imaging with steady-state precession sequence; repetition time msec/echo time msec: 9/5, flip angle: 25°, isotropic resolution: 0.9mm$\times$0.9mm$\times$0.9mm, one signal acquired, bandwidth: 260Hz/pixel). Now that we know where the calibration phantom is in both of these coordinate systems, we deduce the rigid body transformation ($\boldsymbol{H}_{\mathrm{fd}}$) between the patient coordinate system and the coordinate system of the marker frame.

To guide the needle placement procedure we track the needle with an attached marker cluster as in our previous AR guidance setups and graphically represent it in the augmented view. We overlay the needle with a thin blue cylinder and represent the forward extension of the needle

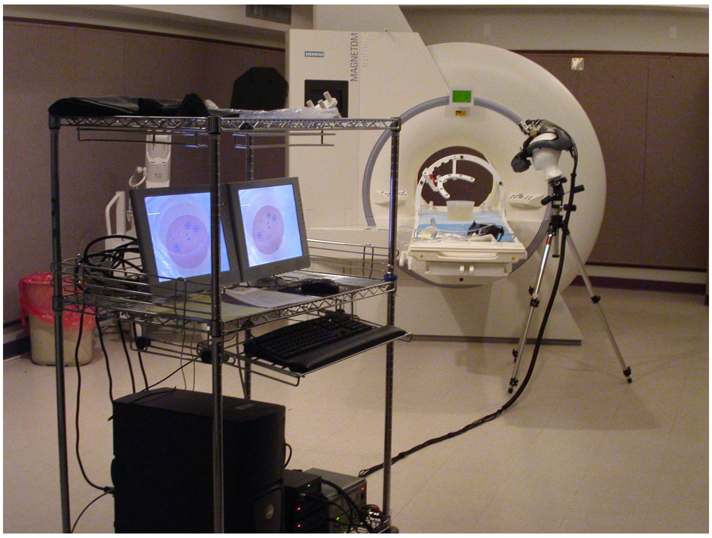



Figure 5.21: *Left:* AR setup for MRI needle placement experiments on gel phantoms. The RAMP setup had to stay in a safe distance from the strong magnetic field of the scanner. *Right:* Dr. Wacker performing an AR guided needle placement. Then Gadolinium gel is injected through a thin, flexible tube through the hollow needle to mark the location of the needle tip inside the phantom.

with a continuation of that cylinder in yellow. The needle extension guides the user to aim the needle toward the target center before the actual insertion.

During the procedure, the AR system visualizes the target centers at their real locations in the augmented view. Contrary to the setup in Chapter 5.3.1 we don't display the whole segmented target structure as a wireframe model. We found that a small disc with a concentric ring gives a very clear spatial impression for the precise target location. To get an improved feedback of the distance between the current location of the needle tip and the target center, we display another concentric ring around the target, with a diameter that is proportional to this distance. Since advancing the needle is usually a forward motion, this type of visualization helps to gain a better *depth perception* along the line-of-sight. To further enhance the spatial perception, we overlay an MRI slice which contains the targets itself.

5.3.4 Phantom Study under MRI-Guidance

We created phantoms by filling plastic containers with Natrosol gel. To simulate lesions we embedded several ring-shaped plastic structures into the gel. All embedded structures were located about 7cm below the surface and were not visible to the user. The gelatin was stiff enough that the needle insertion could not create any significant internal deformations of the phantom.

Figure 5.22: *Left:* "Outside view" of the needle placement procedure. The needle is advanced toward an invisible target inside the gel phantom. *Right:* Augmented view of the needle placement as shown by the AR guidance system. The target structures are represented by discs, and the needle as well as its forward extension are visualized as thin cylinders. The user is provided with a stereoscopic view of the augmented scene.

- The first phantom contained plastic rings with 12mm diameter to show the clinical potential of AR guidance for interventions performed with MRI.
- The second phantom contained plastic rings with 6mm diameter and was used for a more quantitative study.
- A third phantom did not contain any physical targets but was used for needle placements on virtual targets to estimate the *user error*.

For all our tests, we used an MRI-compatible 20cm long 18-gauge needle. Figure 5.21 shows the experimental setup and Dr. Wacker performing an AR guided needle insertion.

Once the phantom has been placed onto the scanner table, underneath the marker frame, we take a set of coronal two-dimensional MRI images using a transverse T1-weighted spin-echo sequence (200/4.4, 90˚flip angle, 5mm section thickness, 256×256 matrix, 0.7mm/pixel in-plane resolution). At the scanner console, we extract the 3D coordinates of the centers of all targets inside the phantom. Those coordinates are transfered to the AR system together with the DICOM images of the MRI scan.

Images of the AR-guided needle placement procedure are shown in Figure 5.22. Outside the bore-hole the user directly advances the needle along a straight path until the graphical representation of the needle tip reaches the center of the target. The *in-situ* visualization of the procedure gives direct and intuitive guidance; the needle placement task can easily be performed within a

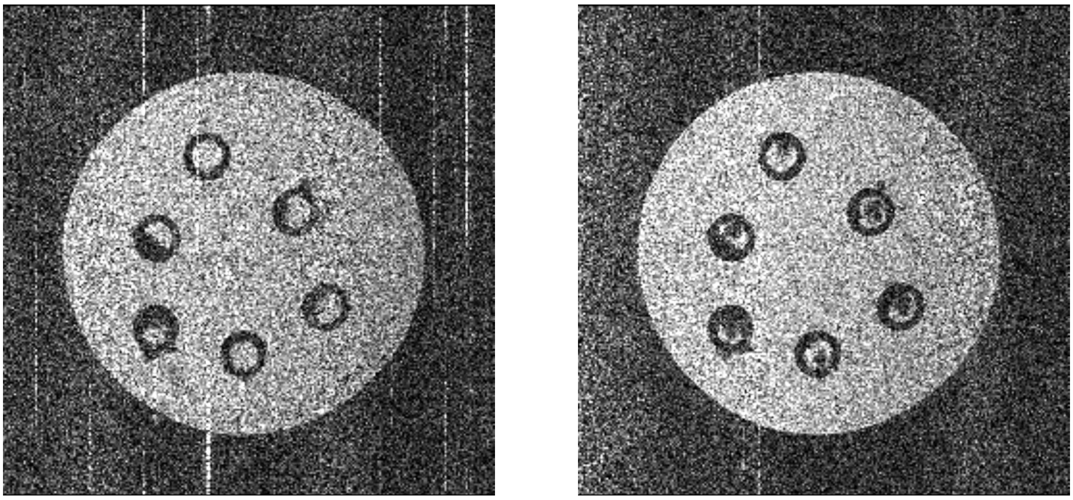


Figure 5.23: *Left:* MRI image of first phantom with six 12mm targets. The targets are embedded about 7cm below the surface of the gel phantom. The same MRI image is shown in the augmented view. *Right:* Control scan of the same phantom after small metal pieces were deposited in an AR guided needle placement procedure.

few seconds. After the placement, we record the position of the needle tip in the marker frame coordinate system and measure its position with a control MRI scan.

We first used the phantom which contained six simulated lesions in form of plastic rings, each with a diameter of 12mm. Under AR guidance, the interventional radiologist placed the needle into each of the targets. Once the needle tip reached the target—according to the guiding graphics—he injected a small metal piece through the hollow needle into the target and retracted the needle to aim for the next target. After all targets were utilized we took a control scan of the phantom to analyze the result. Figure 5.23 displays the MRI scan of the phantom before the procedure on the left, and on the right the control scan that we took afterward. All of the six targets were hit very close to the their centers in the first trial. Each needle placement was performed in less than 10 seconds.

In a second experiment, we used the phantom that contained plastic rings with 6mm diameter. This time, needle placements for 19 targets were performed under AR guidance. To avoid possible movements of the needle during injection of metal pieces, we decided instead to inject small amounts of Gadolinium through a thin flexible plastic tube, which we kept connected to the rear end of the hollow needle throughout the whole procedure. Figure 5.24 displays the MRI scan that we took before the procedure, and the control scan that we took afterward. The mean puncture time from the display of the targets to the final needle position was 4.2 seconds, with a

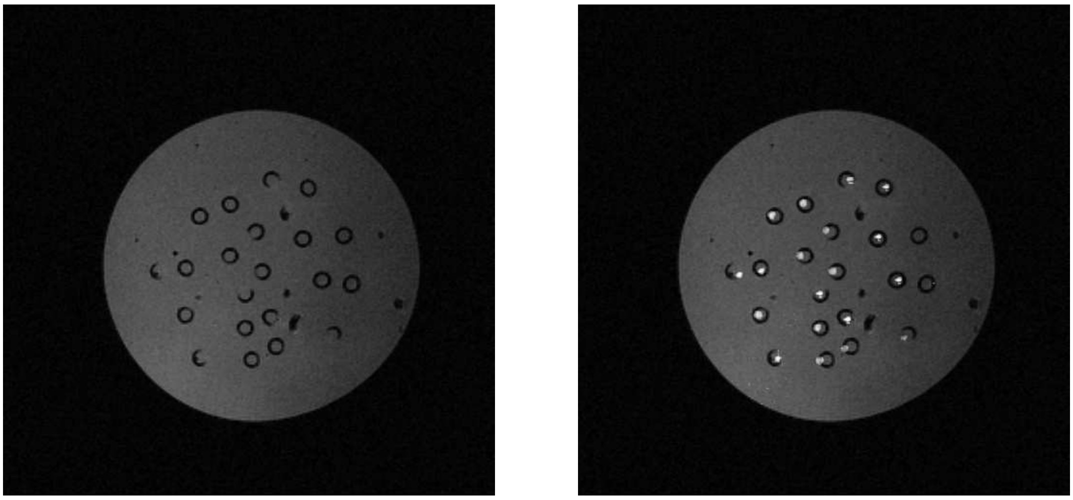


Figure 5.24: *Left:* MRI image of second phantom with 6mm targets. The targets are invisibly embedded about 7cm below the surface of the phantom. The same MRI image is used to create the augmented view. *Right:* Control scan of the same phantom after Gadolinium was injected into the targets in an AR guided needle placement procedure.

range of 1.8–9.3 seconds. All 19 targets were clearly hit in the first trial. The average minimum in-plane distance to the inner edge of the ring, beyond which a needle placement would have been a miss, was 1.44mm. In four cases, the distance was 0mm, which corresponds to a placement of the gadolinium droplet directly at the inner edge of the ring. With regard to the out-of-plane error, it could be seen, that in all experiments the target rings, which had a height of 7mm, were clearly hit in between this vertical range. Both phantom experiments show that the overall error for a stationary target is less than 3mm and that the procedure can be performed in a few seconds.

In a third phantom study, we investigated how well the user can follow the given AR guidance. We recorded the distance between the tip of the virtual needle and the center of the virtual target. This *virtual placement error* was measured in 51 trials, where the virtual targets were placed in a range of 3cm to 10cm below the surface of a Natrosol gel phantom. Figure 5.25 shows the virtual placement error for each of the 51 trials. The maximum error is 2.6mm. The mean error is 1.1mm with a standard deviation of 0.5mm. From these results we conclude that the user is able to follow the AR guidance in our setup with an average error of about 1.1mm. Compared to the 2.5mm average virtual placement error in Chapter 5.3.2, we can derive that a more precise visualization of the actual target center and a pronounced visualization of the virtual distance between needle tip and target center (in shape of a concentric ring of varying diameter, proportional to this distance) gives more direct visual feedback during targeting, leading to a better operator performance.

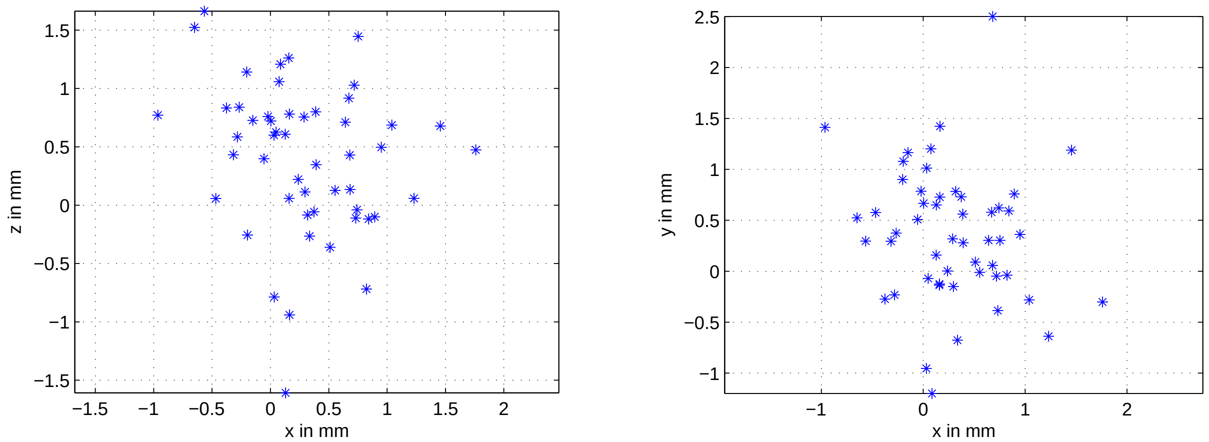


Figure 5.25: Virtual placement error of 51 needle placements. The virtual targets were placed in a range of 3–10cm below the surface of a Natrosol gel phantom. *Left:* In-plane error. x-z-plane is parallel to the scanner table. *Right:* y-coordinate represents the out-of-plane error.

5.3.5 Animal Study under MRI-Guidance

As shown in the previous section, we found that our AR system can provide precise, reliable, and intuitive guidance for needle placement procedures on gel phantoms. As the next step toward clinical applications, we conducted first animal experiments. We performed a set of AR guided needle placement experiments on three anesthetized pigs that weighed 22–40kg each. The experimental protocol was approved by the animal care and use committee at Case Western Reserve University. The animals were anesthetized with an intramuscular injection and an infusion throughout the procedure.

For the experiment one pig at a time was placed on the table of the MRI scanner underneath the marker frame. We chose a supine position (on the back) for one pig and a prone position (on the belly) for the other two pigs. To visualize the anatomy and define target structures we acquired coronal and transverse MRI images and transfered the images via the DICOM connection to our AR system. At the console of the MRI scanner, the interventional radiologist selected target structures for the needle placement procedure. The 3D coordinates of these targets were recorded on the MRI scanner console and transfered to the AR system.

During the AR guided procedure, images and targets were visualized in a similar way as for the phantom experiments. Besides the target, presented as a 5mm disc with a ring, and the size-varying concentric ring demonstrating the target-needle-distance, semitransparent MRI sections containing the target and the skin entry site were augmented into the real view of the animal. This helped to enhance the perception of the target location and to avoid crossing important internal

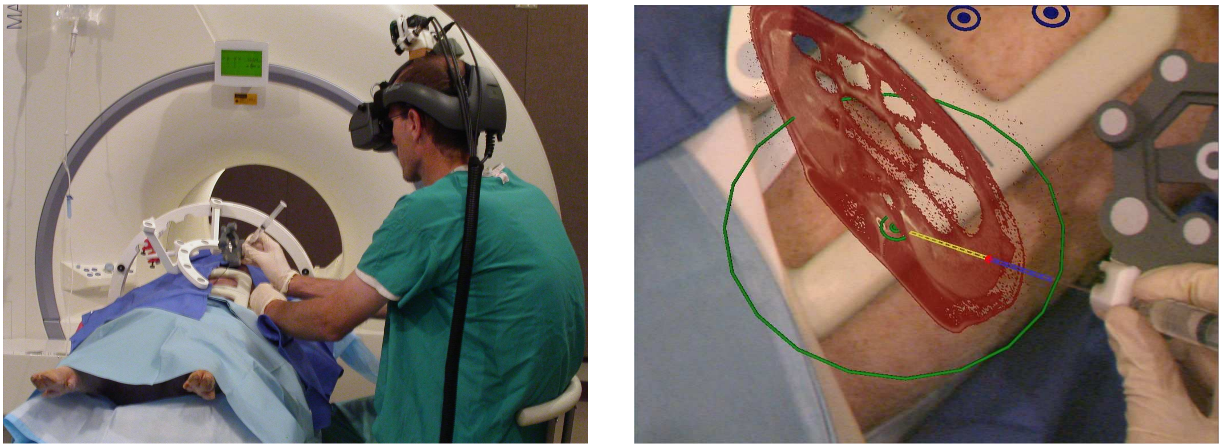


Figure 5.26: *Left:* Dr. Wacker performing an AR-guided needle placement procedure on an anesthetized pig. *Right:* Augmented view as seen through the HMD. An axial MRI slice is visualized *in-situ* together with the marked target structures.

structures during the needle insertion. Ten puncture attempts were performed under AR guidance by Dr. Wacker using

- the pancreatic tail (3 times),
- the gallbladder (3 times),
- a renal calyx (2 times),
- a central bile duct (2 times)

as targets. Figure 5.26 shows the AR-guided needle placement procedure on one of the pigs on the left side, and a snapshot of the augmented view on the right side. To give a better three-dimensional understanding of the AR visualization Figure 5.27 presents two stereoscopic snapshots of the procedure. Immediately after the needle was placed into one of the targets the markerset was clipped off the needle and the animal on the scanner table with inserted needle was moved back into the MRI bore-hole to acquire a control scan and determine the position of the needle tip.

The dynamic stereoscopic augmented view of the animal and the virtual needle extension together with the virtually enhanced target and the augmented semitransparent MRI images provided strong guidance to the performing interventional radiologist during all ten needle placements on the three animals. The average time needed for the procedure, once the needle was placed on the skin of the animal, was 11 seconds (ranging from 3–18 seconds). The average time that the physician needed from the first moment seeing the target in the augmented view to

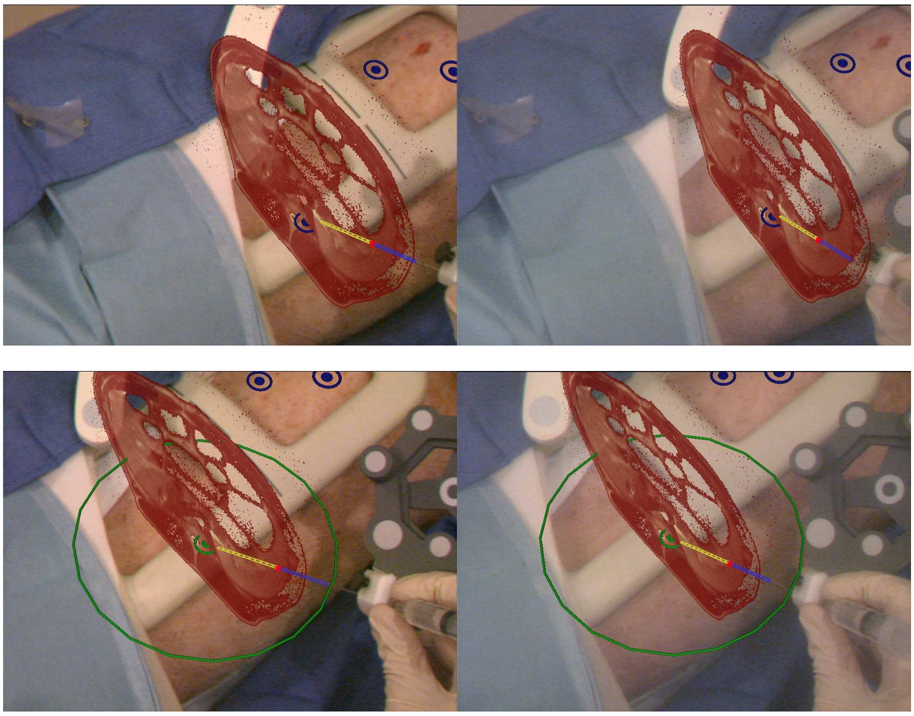


Figure 5.27: AR guidance for needle placements on a pig. *(A "parallel stereo look" at these pictures helps to gain a better understanding of the necessity of the stereoscopic depth cue.)*

reaching the target was 30 seconds. All ten punctures were successful with a maximum error of 10mm in a particular dimension (x,y,z) and an average placement error of 9.6mm with a standard deviation of 4.85mm. It should be noted that the majority of the targets were located more than 10cm below the skin. Figure 5.28 shows MRI images taken before and after the procedure.

The major source for the measured placement error in this animal experiment is motion and, correspondingly, misregistration. Even though the animal was immobilized by anesthesia, the AR system currently cannot account for internal organ motion, such as respiratory motion. The pig exhibited shallow, fast breathing; yet our graphics overlay was a rigid presentation of the MRI scan and the extracted target structures. Furthermore, we noticed that our current markerset assembly limited the workspace range for the intervention surface. Placing the pig onto the scanner table also slightly altered the calibration of the system. For these reasons, we did not expect to achieve the same needle placement precision as in the phantom cases.

However, considering the novelty of this experiment, we were satisfied with the results. The experience of moving the needle through the different tissue structures of the animal while being guided by our AR system was new to the interventional radiologist. He was impressed by the effectiveness, the high performance, and the hand-eye coordination provided by the AR system.

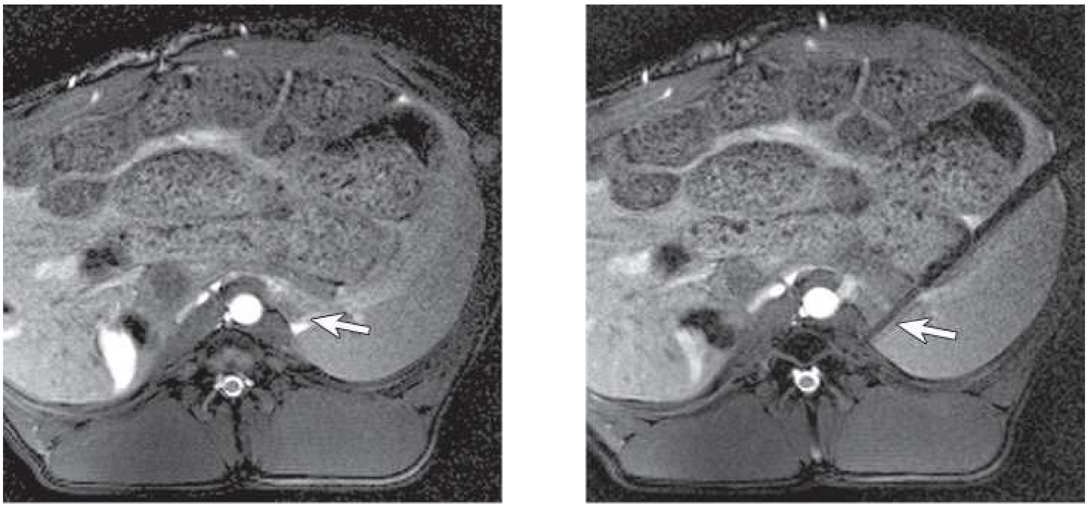


Figure 5.28: Axial MRI images of a pig undergoing an AR-guided needle placement. *Left:* Baseline image used to select the target: the pancreas tail (arrow). *Right:* Control image after AR-guided puncture, with the needle still in position (arrow).

Furthermore, since the procedure is performed outside the bore-hole no spatial constraints for the needle length are given and the operator is not restricted in his choice of an entry point. Robotic systems that operate remote-controlled inside the bore-hole are limited in this respect—and their availability is sparse. The demonstrated AR-guidance has the potential to make the usually tedious and time-consuming MRI-guided interventions in a closed-bore scanner much more efficient. The system is able to give unique, three-dimensional, and intuitive guidance for the needle placement procedure which greatly simplifies and speeds up the targeting process and the intervention itself.

Ultrasound and CT will remain the major tools for monitoring biopsies, due to their cost and availability, and MRI is of importance if a lesion can only be seen in an MRI image or if radiation exposure is critical, such as in children. More studies have to be performed and experience needs to be gained to see which percutaneous interventions benefit most from AR-guidance. An experienced interventionalist will not utilize such an AR-guidance device for simple biopsies, although a less-experienced user might find the AR visualization of targets and critical structures very helpful. For thermal tumor ablations that require positioning of multiple applicators and targeting of multiple lesions, as well as other more complex procedures, AR-guidance can be of great value. It not only reduces procedure time and puncture risk and provides the means for more complete and radical therapy, it can also seamlessly utilize MRI for monitoring thermal lesions and assessing end-organ function. Eventually, the cost of AR-guidance would be negligible compared with reimbursement for thermal ablation procedures and patient benefit.

Figure 5.29: AR visualization of real-time ultrasound images. *Left:* RAMP setup adapted for ultrasound visualization with a tracked hand-held B-probe. *Right:* Augmentation of a breast phantom with its live ultrasound image. The actual US image is naturally the same in left and right view, but its stereoscopic AR visualization puts the slice spatially properly aligned underneath the US probe—inside the imaged object.

5.4 Diagnostic and Interventional Ultrasound

Ultrasonography is a widely utilized imaging technology for diagnostic and interventional purposes (Chapter 2.3.4) due to its low cost, availability, real-time feedback, and avoidance of x-ray radiation by utilization of harmless ultrasound waves. Typical ultrasound transducers acquire 2D cross-sectional images to a certain depth into the imaged body structure. This *B-mode ultrasound* technology is commonly available, but also 3D ultrasound transducers that are able to acquire a subvolume of the imaged body part have been introduced recently. However, compared to CT or MRI images, the quality of ultrasound images is very noisy, is much more difficult to read, and does not provide all the structural information that the other imaging modalities can provide.

As medical imaging technology, ultrasound (US) imaging has been the first modality that was integrated into an early AR system [Baj92, Sta94]. We have studied the integration of ultrasound imaging into the RAMP system for the AR visualization of live ultrasound images in diagnostic and interventional contexts [Kha03a, Sau02a, Sau01b].

5.4.1 AR Visualization of Live US

Unlike CT or MRI imaging information, ultrasound images can be acquired and integrated into the AR view in real-time. The images that the ultrasound probe takes should be displayed in the location where they were taken from. Figure 5.29 shows the RAMP system adapted for in-situ ultrasound visualization and a monoscopic snapshot of an augmented breast phantom. We integrated a Siemens Sonoline Elegra US system with a B-probe 5.0HDPL40 transducer.

The live ultrasound images need to be captured by the RAMP computer in real-time. For this purpose we added another video frame grabber board to the PCI slot of the system. At this point the single PC has to process four video streams with 30Hz in real-time: two color video cameras for the left and the right eye views, the monochrome tracker camera video, and the monochrome ultrasound video stream. To avoid a bottleneck within the bus system of the motherboard, we take advantage of the dual PCI bus architecture, with two video streams channeled through each of them. Since we don't have access directly to the ultrasound datastream of the scanner we utilize its video output, which is originally only connected to the monitor screen of the ultrasound machine, as input to the frame grabber of the RAMP computer. During the augmentation process we upload the incoming US video images into the texture memory on the graphics card (similar to the video streams from the scene cameras) and crop that video stream to the section of the video image that displays the ultrasound image.

The AR system needs to register the live US images to the environment, specifically to the hand-held US probe. Furthermore, the system needs to keep constantly track of the pose of the hand-held US probe in respect to the workspace. The goal is to have the ultrasound image virtually attached rigidly to the hand-held US probe, as if the image was connected to the transducer and "slicing" into the imaged object. For this purpose, we attach a cluster of retroreflective markers to the top part of the US transducer, thereby defining a transducer coordinate system.

To determine the transformation between the image coordinates of the ultrasound image and the transducer coordinate system, a calibration procedure is performed, that utilizes a water-filled phantom with retroreflective markers above and below the water surface. The tracker camera keeps track of the top markers (and the marker cluster on the US probe), while the lower markers are captured with the ultrasound probe and their image coordinates manually extracted. The complete calibration procedure is detailed in [Sau01b, Sau02a]. The determined transformation provides the means to place any incoming ultrasound image spatially underneath the US probe in a place that corresponds to the actual physical object, where the US slice is acquired from. The left side of Figure 5.30 demonstrates this visually inside a phantom test object. A thin water-filled phantom with embedded rubber band pieces is augmented with its live ultrasound scan. This type of augmented reality visualization of a live ultrasound stream permits intuitive access to the acquired imaging information. The hand-eye coordination becomes much more clear than a typical view of the US imaging information on a separate screen.

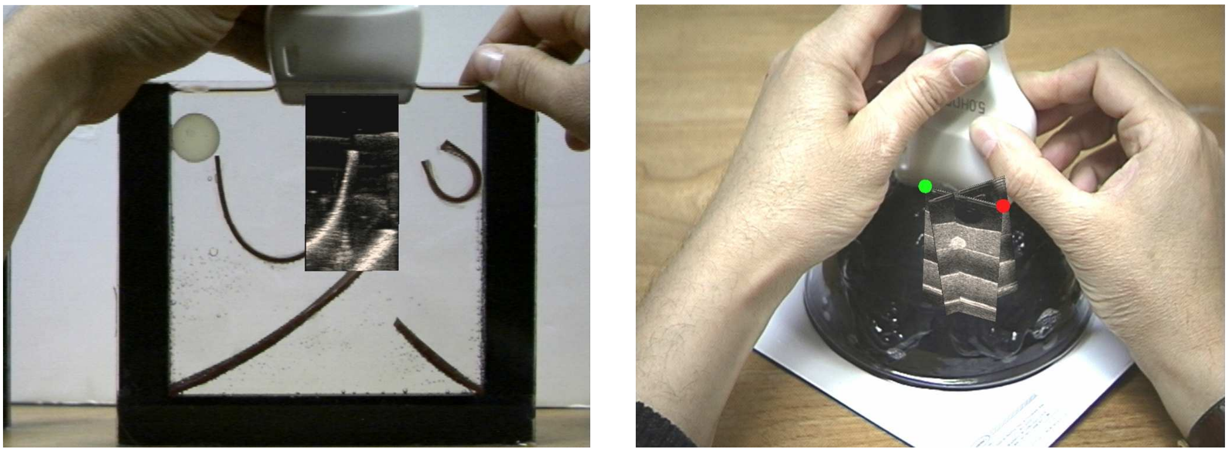


Figure 5.30: *Left:* Ultrasound augmentation of a test phantom with embedded rubber bands. *Right:* Augmentation of a breast phantom with intersecting ultrasound images, one of them frozen in space and time.

5.4.2 AR for US-Guided Needle Biopsies

We developed two methods for AR visualization of ultrasound guided needle biopsies. They differ in the way they present the ultrasound data and provide needle guidance. A (monoscopic) snapshot of the augmented image is presented for each method in Figure 5.31.

- The first implementation [Sau02a] does not utilize instrument tracking but provides guidance through the graphical presentation of an optimal access path. This method is an extension of the previously described in-situ visualization of 2D ultrasound images.

- In our second method [Kha03a] we do not visualize the 2D ultrasound images as-is, but utilize a free-hand sweep and 3D reconstruction method to present the target volume as volume-rendered graphics and provide guidance through needle tracking.

In the first, earlier implementation the needle placement guidance follows this workflow:

1. *Positioning of transducer:* The user looks at the live 2D ultrasound images in-situ and positions the transducer in a way that the target lesion is well visible in the ultrasound slice.

2. *Marking of target in 3D:* Next he marks the target area on the 2D plane of the ultrasound slice with a gaze interface. This is accomplished by a virtual circle in the center of the left augmented view, which the user aligns with the target in the in-situ ultrasound image and then hits a footswitch. Through the intersection of the optical path from the left scene

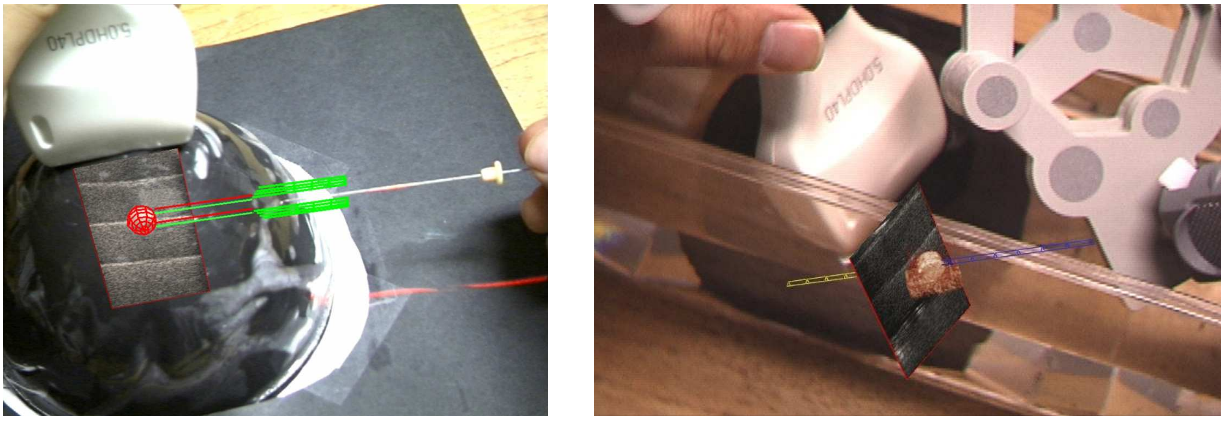


Figure 5.31: AR visualization methods for ultrasound guided needle biopsies in breast phantoms. *Left:* 2D visualization of live ultrasound and guiding graphics for an optimal needle access path. The needle itself is not tracked or represented by graphics. A transducer-attached laser helps to align the needle in the ultrasound plane. *Right:* Visualization of 3D reconstructed ultrasound by volume rendering. The needle is tracked and represented by a blue virtual wireframe cylinder that extends beyond the needle tip in yellow to guide the targeting of the lesion.

camera (through the center of the circle) with the ultrasound plane, the RAMP system estimates the 3D coordinates of the target center and visualizes this with a red sphere, fixed in the coordinate system of the phantom.

3. *Positioning of needle in ultrasound plane:* A red laser, which we attached to the handle of the US transducer, now guides the user to align the needle with the ultrasound plane, by projecting a bright red line to the right side of the transducer. Furthermore, the bright line projects along the side of the breast phantom and therefore provides the user with possible entry points on the phantom surface.

4. *Adjusting in-plane needle angle:* Once again, the user utilizes the gaze interface, by aligning the left camera, which is augmented in the center with a virtual circle, with the chosen entry point on the phantom and presses the footswitch. Now the RAMP system calculates a the line in 3D that connects the target center and the entry point to visualize guiding graphics outside the phantom, which supports the user in aligning the needle with this target line in the ultrasound plane.

5. *Inserting the needle:* Supported by the guiding graphics the user inserts the needle into the phantom, and if done correctly will see the needle appearing into the ultrasound image, aiming for the target, until the target is hit.

This method for ultrasound guided needle biopsies was tested on breast phantoms by several users. The biopsy guidance lead to successful results in almost every attempt. Nevertheless, the workflow of utilizing the gaze interface twice, as well as checking the needle alignment in two different planes simultaneously (laser guide for ultrasound plane and virtual graphics for in-plane guide) throughout the insertion procedure makes this method a bit cumbersome.

The second method of AR visualization for US guided needle biopsies utilizes 3D volume rendering and needle tracking, similar to the guidance approaches in Chapter 5.3 with CT or MRI. A method that compounds free-hand 2D ultrasound scans into a single 3D volume is described in [Kha03a]. Since our transducer is tracked with respect to the workspace, we utilize a slowly performed sweep of the ultrasound probe, covering the target structure, to reconstruct a 3D voxel-based volume from the acquired data. The resulting workflow of the implemented needle biopsy guidance is as follows:

1. *Target localization:* By utilizing the in-situ live 2D ultrasound the user localizes the target lesion and places a 3D bounding box around the region of interest.

2. *3D reconstruction:* The user sweeps the ultrasound probe through the selected bounding box, which is displayed as a stationary wireframe box. Meanwhile the system reconstructs the volume and augments the video cameras with a volume rendered view of the reconstructed part inside the bounding box.

3. *Adjustment of volume rendering properties:* This is an important step since it permits the use of advanced 3D visualization methods, such as volume rendering, and its application-specific optimization of visualization parameters. Therefore, after the volume is scanned completely, the user adjusts the transfer functions to get a better view of the lesion.

4. *Needle insertion:* The needle is inserted into the volume rendered lesion by utilizing a tracked and virtually augmented needle. The graphical extension of the needle guides the user into the target, similar to our MRI and CT needle placement studies.

We performed this guided needle biopsy on four different high-density gel lesions embedded in a breast biopsy phantom. All lesions had a diameter in the range of 7–10mm. After each needle placement we reconstructed another volume with the inserted needle to confirm the success of the experiment. Since the ultrasound images and reconstructed volumes exhibit the typical noise of ultrasound scans, it is not possible to quantitatively analyze the result by measuring the distance between needle tip and lesion center as we did in the CT and MRI guided studies. However the rendered post-op scans show that all lesions were hit successfully. Figure 5.32 shows snapshots

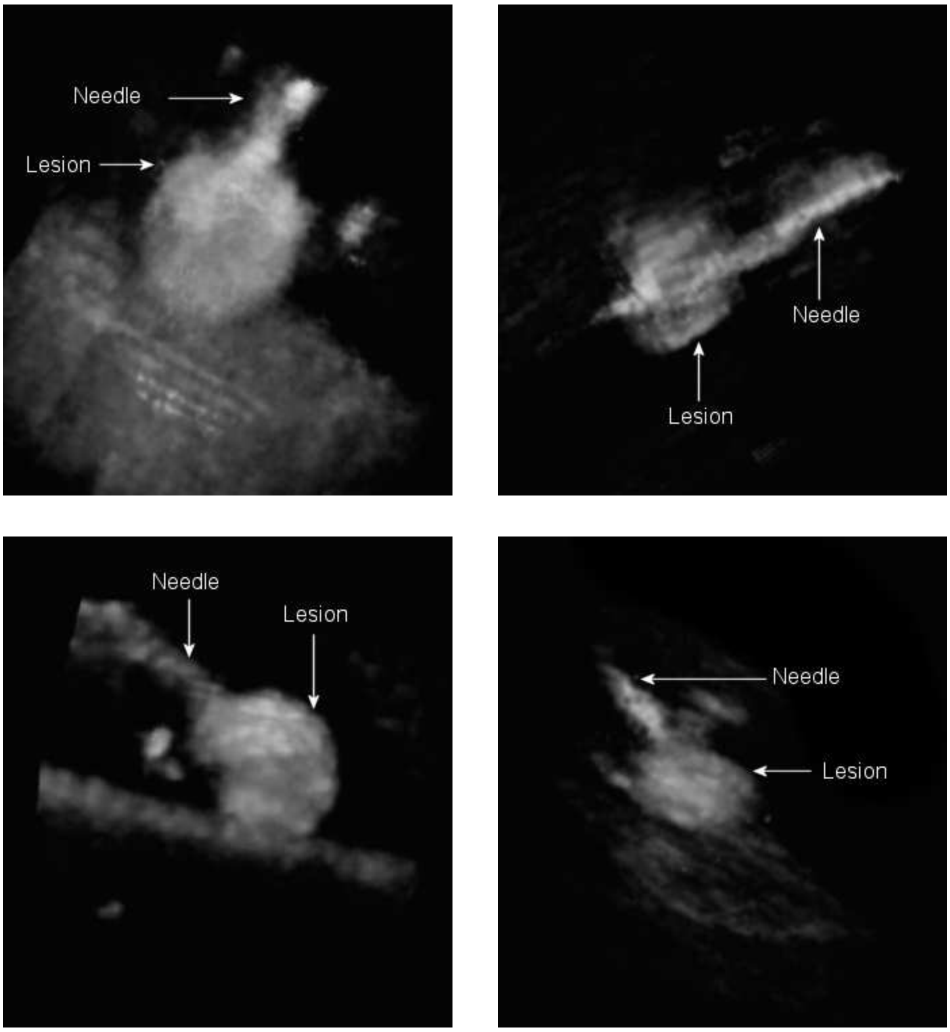


Figure 5.32: 3D reconstructed control scans after AR guided needle placements with ultrasound. All four cases show a successful placement of the needle into the 7–10mm target lesions.

of four post-op scans. Although the users were all novices in needle placement tasks, the augmented reality setup provided intuitive guidance and resulted in repeatedly successful needle placements performed in less then 10 seconds each. It is the strength of this second guidance approach that builds atop a novel unified augmented reality ultrasound platform, integrating the acquisition of ultrasound B-plane slices from around the target location with the reconstruction of a local 3D volume dataset and its volume-rendered in-situ visualization to support the assisted needle placement procedure.

Chapter 6

Future Work

Augmented reality as a means for interventional image guidance faces many challenges, but also shows strong potential to tackle difficult interventional scenarios, as studied in the previous chapters. Even though the introduced RAMP prototype system has been brought close to clinical settings in potentially benefiting applications, further work needs to be done to realize a breakthrough into the interventional arena as a product. Future work should focus on three core topics:

- *Registration and tracking:* The current challenges are mainly patient and organ motion during the intervention, especially due to respiratory motion. The continuous update of registration parameters might need to go beyond the currently tracked ones.

- *Visualization:* The miniaturization of the head-mounted video see-through display is here of primary interest. Furthermore, perceptual issues concerning the visual relationship between the surfaces of real and virtual objects need to be studied to provide optimized graphics overlays.

- *Applications:* More interventional applications, beyond the introduced ones, need to be investigated that make it worth-while to consider the extra amount of money and possibly discomfort of an augmented reality guidance device.

As has been shown in the reported studies, respiratory motion is of concern when utilizing pre-operative images as a basis for the guiding graphics during the intervention. A first study is discussed in [Kha04], which utilizes optical markers attached to the patient's abdomen to generate a patient-specific model of the marker motion in respect to the motion of, for instance, a liver lesion throughout the breathing cycle. For that purpose the breathing cycle needs to be monitored and the target region imaged throughout the cycle prior to the intervention. Also, an appropriate location for the optical markers on the abdomen needs to be determined to gain optimal results. The major drawback of this approach is that additional markers might clutter the field-of-view of the physician and slight motions of the abdominal area might introduce big errors into the chosen model of the target motion.

Another approach to deal with respiratory motion is by utilizing real-time interventional imaging modalities, such as ultrasound, to track internal organ deformations. The system could for instance give interventional image guidance on the MRI scanner table, by utilizing pre-operative MRI images that are registered to real-time ultrasound images throughout the intervention and accordingly modified. This method puts high demands on the precision and speed of the chosen inter-modality image registration technique. The advantage is that the ultrasound images itself are not used to perform the guidance, but high-quality MRI images can be utilized. Nevertheless, the ultrasound information could be sufficient to track internal organ motion and deformations.

As has been shown in most studies performed with the RAMP system, the best guidance results are achieved when target and instrument are represented by virtual graphics. This is a pure virtual–virtual interaction, although the real part of the image is used for an overall orientation. Nevertheless, it is of future importance to research how in-situ visualizations of deep-lying, disconnected internal structures can be easily related and perceived in respect to the surface of the real object. Shadows, lighting, eye accommodation, and some of the remaining psychological and physiological depth cues (see Chapter 2.4.3) will need to be studied here to realize an optimal scene fusion. However, it should be noted that the stereoscopic and kinetic depth cues, as well as the zero time-lag between graphics and video, provided by the RAMP system, already realize a very believable experience of a single fused scene to the user, which the illustrations in this dissertation cannot easily represent. Isolated virtual structures that are visualized deep inside the real object, without a connection to the surface, remain challenging.

It is of primary interest to explore and study more interventional applications for AR. In [Sau06] the RAMP system is brought into an orthopedic scenario. Here a thoracoscope is utilized to perform minimally invasive surgery through the thorax at the front of the spine. The RAMP system could potentially guide this type of surgery by fusing internal images from the thoracoscope and pre-op CT images with the real video images. First results are promising.

Research at the RAMP system is carried on especially at the Chair for Computer Aided Medical Procedures & Augmented Reality at Technische Universität Munich [Hei06, Sie06, Sie04b, Sie04a]. The system has already been adapted to new applications, such as a birth simulator, where in-situ visualization with this system increases the efficiency of the training and simplifies complicated procedures such as forceps delivery.

Chapter 7

Summary

Interventional image guidance has been a research topic since the 1980s due to its potential for enabling minimally invasive surgery with fewer complications. In its beginning, stereotactic neurosurgery utilized head-attached frames, but advances towards frameless devices brought interventional image guidance to a broader application spectrum. The main contribution of this work is the design, development, and study of a new apparatus and approach for interventional image guidance based on augmented reality visualization. AR navigation augments the view of the actual patient with medical images and guidance information.

AR systems rely on three enabling technologies—visualization, tracking, and registration—and face common challenges with respect to real-time operation, display technologies, accuracy and latency, range of operation, and the acceptance by its potential users. Medical imaging is the fourth enabling technology for AR in the interventional field.

With limitations, medical imaging devices can be utilized for interventional image guidance on their own or they can be utilized as image source for sophisticated interventional guidance devices. X-ray imaging, fluoroscopy, CT and MRI imaging, as well as ultrasonography and endoscopy-type imaging provide patient-specific information to plan an intervention and extract target coordinates for guidance. In a continuous acquisition mode all devices could provide interventional guidance, but each modality has its essential shortcomings—with respect to radiation, image quality, real-time acquisition, and flexibility.

Visualization of medical images is commonly performed on a monitor screen utilizing techniques that transform the acquired data into gray-scale or color information. For the purpose of AR visualization, overlaying the rendered medical images onto the real view can be achieved by optical or video see-through techniques, where the first attracts through simplicity and the latter with respect to synchronization and flexibility. Depth cues based on binocular and motion parallax should be utilized to provide proper depth perception. Depending on the approach, AR systems use head-attached, hand-held, or spatial displays for visualization.

Registration of real and virtual views is needed to place augmented graphics in its proper 3D location of the real world at any given point in time. It can be solved through two sub-tasks: the 3D registration of virtual with real objects and the calibration of the real viewing parameters. On the one hand, the calibration task is needed for modeling the virtual cameras to render virtual objects with camera parameters that equal the real camera parameters. On the other hand, the 3D registration task needs to resolve the correspondence between medical datasets and anatomical features of the actual patient. An AR system has to minimize static and dynamic errors to create a believable AR experience.

Tracking needs to be involved to update parameters of the registration that change over time. Typically the pose of the user and therefore the pose of head-mounted see-through displays and hand-held instruments change, whereas intrinsic camera parameters and the 3D registration parameters stay constant. Since high update rates are required, dedicated tracking systems based on optical and electromagnetic technologies are desired.

Existing approaches for interventional AR guidance can be classified into optical microscope systems, video see-through systems, large semi-transparent screens, tomographic overlay systems, and video endoscope approaches. So far, each reported system is a research prototype and targeted towards one or several specific applications, covering neurosurgery and otolaryngology, cranio- and maxillofacial surgery, needle biopsies and tumor ablations, orthopedics, and cardiovascular and thoracic surgery.

We introduce a novel AR system for interventional image guidance: *RAMP—Reality Augmentation for Medical Procedures*. The video see-through approach exhibits complete control over the visualization of the merged scene, the ability to mix video and graphics in a synchronized and precise fashion, as well as reliability and the possibility of sharing the AR view—attributes which target the high expectations on an interventional device. We mounted two miniature color cameras and a dedicated tracker camera with an infrared flash rigidly on top of a head-mounted virtual reality display. All cameras are genlocked and connected together with the stereoscopic HMD to a single PC, providing 30 frames/second of real-time video AR images with 1024×768 resolution for each eye. A C++ and XML based component- and object-oriented software architecture provides the means to adapt the system to varying hardware setups and to prototype a variety of applications. The stereoscopic blending of video and graphics is implemented in OpenGL with virtual stereo cameras that utilize registration and real-time tracking parameters of the actual head-mounted cameras.

Tracking is realized with the single monochromatic head-mounted camera. Inside–out tracking helps with the perceived accuracy of the graphics overlay and offers advantages over of-the-shelf outside–in optical tracking systems that are based on multiple cameras and triangulation techniques. A set of retro-reflective markers in the periphery of the workspace and instrument-attached marker clusters are utilized for the real-time estimation of the poses of the user's head and of hand-held instruments. The 2D marker centers are extracted with sub-pixel precision from the image moments of the infrared camera image. A special calibration technique is used to estimate the intrinsic and relative extrinsic parameters of the head-mounted camera triplet with a dedicated calibration phantom offline. Radial distortion of the fish-eye lens of the tracker camera is properly modeled up to the sixth order. Real-time pose estimation of the user's head and of hand-held instruments is based on a unique scheme that extracts three characteristic markers as input for an initial 3-point algorithm and, furthermore, invokes a non-linear optimization of the reprojection errors of all markers to refine the poses, altogether in less than 10ms.

Sub-pixel precision for calibration and tracking is essential because of the use of a fish-eye lens; small errors here translate into bigger overlay errors in the scene cameras. Regarding head-tracking, the graphics is perfectly aligned with stationary objects in the scene and does not exhibit any visible jitter. For instrument tracking, the marker cluster design determines how much image noise amplifies into a jitter of its estimated pose and ultimately into jitter of the graphical representation of the instrument in the AR view. By studying the tracking error of a family of marker clusters, based on many Monte Carlo simulations and several real experiments with varying marker cluster parameters, we optimized the cluster design, favoring a non-coplanar set of 8 markers with a cluster radius of 45mm. Tracking a needle with that cluster resulted in a jitter of less than 1mm of the tip position—an invisible jitter in the AR image.

We analyzed the potential of the system for AR guided navigation. More than one hundred users performed a needle placement procedure on a phantom box with hidden 6mm targets with a success rate of 100%. This user interface was experienced as very intuitive and might be especially helpful in a setting with complex hidden anatomy. In another user study we investigated how this AR navigation method performs in comparison with three types of stationary, monitor-based navigation methods. The results show that stereoscopic 3D visualization with a dynamic viewpoint contributes much to the overall success and efficiency of the procedure.

Together with surgeons and radiologists we studied the RAMP system in several medical scenarios. For surgery planning we developed an interactive setup that allows the surgeon to efficiently explore volumetric datasets, such as CT or MRI scans of the head, and plan an intervention more intuitively, in-situ. The system utilizes volume rendering and multi-planar reformatting techniques in combination with a tracked hand-held input device for direct interactions. This setup has been exhibited at the European Congress of Radiology and collected positive feedback about its clinical potential from a wide spectrum of physicians.

We adapted the system to the neurosurgical MRI operating suite at UCLA by attaching the marker frame to the head clamp of the scanner table and utilizing a dedicated calibration phantom to transform scanner coordinates into the marker frame coordinate system. The neurosurgeon's task was to extract tumor tissue under AR navigation. The pre-clinical setup utilized watermelon and mashed potato phantoms to study its effectiveness. Without instrument tracking the neurosurgeon was able to easily reach hidden shallow targets, 1–3cm beneath the surface. Deeper lying structures required tracking and visualization of the instrument. Furthermore, the successful removal of all simulated tumor tissue was confirmed by post-operative MRI scans. At the Hospital of the University of Pennsylvania we could assess the feedback of a neurosurgeon for AR visualization preceding an actual craniotomy.

In interventional radiology minimally invasive needle placement procedures are a crucial method for abdominal tumor ablations and biopsies. At Brigham and Womens Hospital we utilized CT images to perform AR-guided needle procedures on an abdominal phantom. The interventional radiologist was targeting simulated liver lesions of 1cm diameter with an average procedure time of 10 seconds. A post-operative CT scan confirmed the success. We analyzed three types of errors and concluded that 10mm lesions can easily be targeted under AR guidance.

In an MRI suite at the University Hospitals of Cleveland we successfully continued phantom experiments with targets of 12mm and 6mm diameter. We determined that the user is able to follow the AR guidance with an average error of about 1.1mm. First animal experiments were performed on three anesthetized pigs with targets 10cm below the animal skin, which were hit with an average placement error of 9.6mm. Most of the error is caused by respiratory motion and therefore misregistration. However, moving the needle through the different tissue structures of the animal while being guided by AR visualization was novel to the radiology community and initiated much interest in the technology.

We also studied AR visualization of diagnostic ultrasound in a laboratory setting—providing intuitive access to hidden anatomy. We furthermore explored the use of interventional ultrasound for AR guided needle placement procedures on breast phantoms. Also here, a tracked needle and in-situ ultrasound images significantly ease the task of breast needle biopsies.

The presented work encourages the continued exploration of AR technology, moving further towards clinical studies. AR systems provide access to medical images that is compellingly direct and intuitive, and they can help to perform interventions more efficiently and precisely.

Appendix A

Mathematical Symbols

This appendix lists the mathematical notations that were used in this dissertation.

- Scalar values are denoted by small non-bold italic letters, such as x.
- Vectors are denoted by small bold italic letters, such as $\boldsymbol{x}$.
- Matrices are denoted by capital bold italic letters, such as $\boldsymbol{X}$.
- Sets are denoted by capital non-bold italic letters, such as X.
- The inverse inverse of a matrix $\boldsymbol{X}$ is denoted by $\boldsymbol{X}^{-1}$.
- A homogeneous vector is identified by a tilde, such as $\tilde{\boldsymbol{x}}$.
- The equation $x \cong y$ indicates that x and y are equal up to a scale factor.
- The Euclidean norm of a vector $\boldsymbol{x}$ is denoted by $\|\boldsymbol{x}\|$.

Appendix B

German Title and Abstract

Erweiterte Realität in Echtzeit für bildgeführte Interventionen

Das Ziel dieser Dissertation ist es einen neuen Ansatz für bildgeführte Interventionen, basierend auf erweiterter Realitätsvisualisierung – der Vereinigung von realen und virtuellen Bildern, zu entwerfen, zu entwickeln, und zu studieren. Obgleich rahmenlose stereotaktische Navigationssysteme, die kommerziell verfügbar sind, momentan bildgeführte Interventionen unterstützen, sind Chirurgen und interventionelle Radiologen darauf beschränkt auf einen Bildschirm zu sehen, weg von dem eigentlichen Patienten, was die Hand-Auge-Koordination entscheidend kompliziert. Für einen fundamentell neuen Ansatz, aufbauend auf erweiterter Realität, ist eine geeignete Kombination von dedizierten Beiträgen zu mehreren Ingenieursdisziplinen notwendig, insbesondere in Systementwicklung, Rechnersehen, Datenvisualisierung und Mensch-Computer Interfacedesign. Zusätzlich müssen Beiträge in den Medizinwissenschaften geleistet werden, die auf vorklinischen Studien und Auswertungen des neuen Ansatzes beruhen. Wichtige Ergebnisse in diesem Bereich dienen wiederum konstruktiv der Weiterentwicklung des Erweiterten-Realitäts-Navigationssystems.

Beiträge auf dem Gebiet der Systementwicklung. Die Komplexität eines Systems für erweiterte Realität kann nicht mit der Summe seiner Teile gleichgesetzt werden. Diese Arbeit gibt

deshalb einen breiten Überblick und kategorisiert alle publizierten Prototypen auf dem Gebiet der erweiterten Realität für interventionelle Bildführung und dient als die erste umfassende Referenz auf diesem Gebiet. Ein neuartiges Setup zur erweiterten Realität wird eingeführt, erstellt und optimiert für interventionelle Navigationsaufgaben, das voroperative Bildgebung, speziell CT und MRI, sowie intraoperativen Ultraschall einbindet. Aufbauend auf einer einzelnen PC-Platform mit vier unabhängigen Video-Digitalisierkarten und synchronisierten Kameras kombiniert dieses System einzigartig kopf-basierte stereoscopische Videoaufnahme und Visualisierung mit optischem Tracking von Kopf und Instrumenten in einem sehr effizienten aber zugleich modularen „closed-loop" Ansatz. Aufgrund seiner besonderen Architektur bietet das System eine bisher unerreichte Leistung und Genauigkeit von hochauflösender erweiterter Realität in Echtzeit und ermöglicht dadurch komplizierteste Navigationsaufgaben.

Beiträge auf dem Gebiet des Rechnersehens. Diese Arbeit entwickelt Methoden zur Multi-Camera-Kalibrierung und zum Tracking in Echtzeit. Ein neuartiges kopf-montiertes Kameratriplett wird beschrieben, bei dem zwei Kameras den stereoscopischen Blick der Szene aufnehmen, während eine dritte, nah-infrarote Kamera mit einem Weitwinkelobjektiv und einem synchronisiert pulsierendem Infrarotblitz zu Pose-Schätzungen in Echtzeit, durch Tracking von retroreflektierenden Markern in der Szene, genutzt wird. Eine spezialisierte Kalibrierungstechnik, mit einem dedizierten Kalibrierungsobjekt, wird zur Kalibrierung aller intrinsischen und relativ-extrinsischen Parameter des Kameratripletts entwickelt, die die sehr unterschiedlichen optischen Eigenschaften der drei Kameras berücksichtigt. Diese Technik baut auf etablierten photogrammetrischen Methoden zur Kalibrierung einer einzelnen Kamera mit Hilfe von 2D–3D Punktkorrespondenzen auf. In Bezug auf das optische Tracking in Echtzeit wird ein umfassender Ansatz zur Pose-Schätzung entwickelt, der ein optimiertes Klusterdesign von retroflektierenden Markern, eine Strategie zur Markeridentifizierung in den Bildern der Trackerkamera und eine Methode zur Schätzung der extrinsischen Kameraparameter, basierend auf einem intitialen 3-Punkt Algorithmus und nichtlinearen Parameteroptimierungen unter Nutzung aller verfügbaren 2D–3D Punktkorrespondenzen, beruht. Dieser umfassende Ansatz erreicht genaueste Pose-Schätzungen des Kameratripletts in Bezug zu stationären Markerrahmen und Markerklustern, die an handgehaltenen Werkzeugen befestigt sind, in weniger als 10 ms. Die Robustheit der Trackingergebnisse wird in Monte-Carlo Simulationen und in realen Experimenten ausgewertet.

Beiträge auf den Gebieten der Datenvisualisierung und des Mensch-Computer Interfacedesigns. Das vorgestellte System zur erweiterten Realität bietet einen komplett neuen intuitiven Zugang zu dreidimensionalen Datensätzen in direktem Bezug zu echten Objekten und Interaktionen. Die einzigartige Kombination von extrem schnellen Rendering volumetrischer Datensätze (wie CT-Scans oder Ultraschallscans), kopf-montierter Visualisierung und Kopftracking bietet dem Nutzer Bewegungsfreiraum und ermöglicht es ihm den scheinbar stationären Datensatz aus einer Vielzahl von Blickrichtungen zu studieren. Im Vergleich zu gewöhnlichen, bildschirm-basierten Visualisierungsmethoden, bieten die stereoscopischen und kinetischen Tiefeneindrücke des vorgestellten kopf-montierten Ansatzes einen eindeutig höheren Grad an Immersion – und damit Verständnis – der Daten. Ausserdem bietet das System einen stereoscopischen Blick auf die reale Szene (z.B. den Körper des Patienten) in direktem Bezug zu den virtuellen Daten an, wodurch reale Objekte im Wesentlichen transparent werden, wenn der virtuelle Datensatz ansonsten unsichtbare interne Details darstellt. Interaktionen mit dem Datensatz, die in einer bildschirm-basierten Lösung mittels Mausbewegungen und -Clicks realisiert werden müssen, werden mit Hilfe von getrackten Instrumenten in einer Art und Weise erwirklicht, die Interaktionen mit realen Objekten gleichen. Bespielsweise ist ein getrackter handgehaltener Zeiger ausreichend um beliebige Querschnitte durch den 3D Datensatz zu erzeugen – so einfach wie das Bewegen einer Klarsichtfolie durch den Datensatz.

Beiträge zu Anwendungen von erweiterter Realität. Diese Arbeit leistet überzeugende Beiträge zu potenziellen Anwendungen von erweiterter Realität im Beriech der Intervention. Studien werden durchgeführt, die zeigen, dass dreidimensionale Grafiken, die mittels stereoscopischer und kinetischer Tiefeneindrücke im Inneren des eigentlichen Patienten dargestellt werden, einen sehr fesselnden und direkten Zugang zu anatomischen Informationen zur Planung und Führung einer effizienten Intervention bieten. Chirurgen profitieren speziell bei minimalinvasiven Eingriffen von dieser direkten Überlagerung von medizinischen Bildern auf dem eigentlichen Patienten. Zusätzlich wird die Navigation durch getrackte Instrumente unterstützt, speziell Biopsie- und Ablationsnadeln, die in Form von virtuellen Objekten innerhalb des Patienten visualisiert werden. Unter Nutzung des vorgestellten Systems zur erweiterten Realität werden Ergebnisse von vorklinischen Phantom- und Tierstudien dargelegt, die den Nutzen der Verfügbarkeit von voroperativen Bildinformationen (z.B. CT oder MRI) für den interventionellen Radiologen während des eigentlichen Eingriffes auswerten. Eine erhöhte Genauigkeit der Biopsie, eine Verkürzung der Operationszeit und der Verzicht auf andauernde Röntgenstrahlenbelastung durch Flouroscopie stellen hierbei die Hauptbeiträge dar.

Bibliography

[Ant00] Howard Anton and Chriss Rorres. *Elementary Linear Algebra: Applications Version*. John Wiley & Sons, 8 edition, 2000.

[A.R02] A.R.T. GmbH, 2002. A.R.T. Infrared Tracking System, www.ar-tracking.de.

[Aza04] Fred S. Azar, Nathalie Perrin, Ali Khamene, Sebastian Vogt, and Frank Sauer. User performance analysis of different image-based navigation systems for needle placement procedures. In *Proceedings of SPIE's Conference of Medical Imaging 2004: Visualization, Image-Guided Procedures, and Display*, volume 5367, pages 110–121, February 2004.

[Azu97] Ronald T. Azuma. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, 6(4):355–385, August 1997.

[Azu01] Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. Recent advances in augmented reality. *IEEE Computer Graphics and Applications*, 21(6):34–47, 2001.

[Baj92] Michael Bajura, Henry Fuchs, and Ryutarou Ohbuchi. Merging virtual objects with the real world: Seeing ultrasound imagery within the patient. *ACM SIGGRAPH Computer Graphics*, 26(2):203–210, July 1992.

[Bal82] Dana H. Ballard and Christopher M. Brown. *Computer Vision*. Prentice Hall, 1982.

[Bau01] Martin Bauer, Bernd Bruegge, Gudrun Klinker, Asa MacWilliams, Thomas Reicher, Stefan Riss, Christian Sandor, and Martin Wagner. Design of a component–based augmented reality framework. In *Proceedings of the IEEE and ACM International Symposium on Augmented Reality (ISAR)*, pages 45–54, October 2001.

[Bil02] Mark Billinghurst et al., 2002. ARToolKit computer vision tracking libraries, www.hitl.washington.edu/artoolkit/.

[Bim05] Oliver Bimber and Ramesh Raskar. *Spatial Augmented Reality: Merging Real and Virtual Worlds*. A K Peters, Wellesley, MA, 2005.

[Bir00a] Wolfgang Birkfellner. *Tracking Systems in Surgical Navigation*. PhD thesis, Department of Biomedical Engineering and Physics, General Hospital, University of Vienna, Waehringer Guertel 18–20, 1090 Vienna, Austria, 2000.

[Bir00b] Wolfgang Birkfellner, Michael Figl, Klaus Huber, Franz Watzinger, Felix Wanschitz, Rudolf Hanel, Arne Wagner, Dietmar Rafolt, Rolf Ewers, and Helmar Bergmann. The varioscope AR – a head-mounted operating microscope for augmented reality. In *Proceedings of the 3rd International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 869–877, October 2000.

[Bir00c] Wolfgang Birkfellner, Klaus Huber, Franz Watzinger, Michael Figl, Felix Wanschitz, Rudolf Hanel, Dietmar Rafolt, Rolf Ewers, and Helmar Bergmann. Development of the Varioscope AR, a see-through HMD for computer-aided surgery. In *Proceedings of the IEEE and ACM International Symposium on Augmented Reality (ISAR)*, pages 54–59, October 2000.

[Bir01] Wolfgang Birkfellner, Michael Figl, Klaus Huber, Johann Hummel, Rudolf A. Hanel, Peter Homolka, Franz Watzinger, Felix Wanschitz, Rolf Ewers, and Helmar Bergmann. Calibration of projection parameters in the varioscope AR, a head-mounted display for augmented-reality visualization in image-guided therapy. In *Proceedings of SPIE's Conference of Medical Imaging 2001: Visualization, Display, and Image-Guided Procedures*, volume 4319, pages 471–480, February 2001.

[Bir02] Wolfgang Birkfellner, Michael Figl, Klaus Huber, Franz Watzinger, Felix Wanschitz, Johann Hummel, Rudolf Hanel, Wolfgang Greimel, Peter Homolka, Rolf Ewers, and Helmar Bergmann. A head-mounted operating binocular for augmented reality visualization in medicine–design and initial evaluation. *IEEE Transactions on Medical Imaging*, 21(8):991–997, August 2002.

[Bla95] M. Blackwell, R. V. O'Toole, F. Morgan, and L. Gregor. Performance and accuracy experiments with 3D and 2D image overlay systems. In *Proc. Medical Robotics and Computer Assisted Surgery (MRCAS)*, pages 312–317, November 1995.

[Bla98a] Mike Blackwell, F. Morgan, and Anthony M. Di Gioia, III. Augmented reality and its future in orthopaedics. *Clinical Orthopaedic and Related Research*, 345:111–122, September 1998.

[Bla98b] Mike Blackwell, Constantinos Nikou, Anthony M. DiGioia, and Takeo Kanade. An image overlay system for medical data visualization. In *Proceedings of the 1st International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 232–240, October 1998.

[Bla00] Mike Blackwell, Constantinos Nikou, Anthony M. Di Gioia, III, and Takeo Kanade. An image overlay system for medical data visualization. *Medical Image Analysis*, 4(1), 2000.

[Cha05] Wilson M. Chang, Michael B. Horowitz, and George D. Stetten. Intuitive intraoperative ultrasound guidance using the sonic flashlight: A novel ultrasound display system. *Neurosurgery*, 56(4):434–437, April 2005.

[CM04] Ève Coste-Manière, Louaï Adhami, Fabien Mourgues, and Olivier Bantiche. Optimal planning of robotically assisted heart surgery: Transfer precision in the operating room. *The International Journal of Robotics Research*, 23(4):539–548, May 2004.

[Das06] Marco Das, Frank Sauer, U. Joseph Schoepf, Ali Khamene, Sebastian K. Vogt, Stefan Schaller, Ron Kikinis, Eric van Sonnenberg, and Stuart G. Silverman. Augmented reality visualization for CT-guided interventions: System description, feasibility, and initial evaluation in an abdominal phantom. *Radiology*, 240(1):230–235, July 2006.

[De 01] Stijn De Buck, Johan Van Cleynenbreugel, Indra Geys, Thomas Koninckx, Philippe R. Koninck, and Paul Suetens. A system to support laparoscopic surgery by augmented reality visualization. In *Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 2208 of *Lecture Notes in Computer Science*, pages 691–698. Springer-Verlag, October 2001.

[Dey00] Damini Dey, Piotr J. Slomka, David G. Gobbi, and Terry M. Peters. Mixed reality merging of endoscopic images and 3-D surfaces. In *Proceedings of the 3rd International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 797–803, October 2000.

[Edw95a] P. J. Edwards, D. J. Hawkes, D. L. G. Hill, D. Jewell, R. Spink, A. J. Strong, and M. J. Gleeson. Augmentation of reality in the stereo operating microscope for otolaryngology and neurosurgical guidance. *Computer Assisted Surgery*, 1(3):172–178, 1995.

[Edw95b] P. J. Edwards, D. L. G. Hill, D. J. Hawkes, R. Spink, A. C. F. Colchester, A. Strong, and M. Gleeson. Neurosurgical guidance using the stereo microscope. In *Proceedings of Computer Vision, Virtual Reality, and Robotics in Medicine '95 (CVRMed)*, pages 555–564, April 1995.

[Edw99] Philip J. Edwards, Andrew P. King, Calvin R. Maurer, Jr., Darryl A. de Cunha, David J. Hawkes, Derek L. G. Hill, Ron P. Gaston, Michael R. Fenlon, Subhash Chandra, Anthony J. Strong, Christopher L. Chandler, Aurelia Richards, and Michael J. Gleeson. Design and evaluation of a system for microscope-assisted guided interventions (MAGI). In *Proceedings of the 2nd International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 1679 of *Lecture Notes in Computer Science*, pages 843–851. Springer-Verlag, September 1999.

[Edw00] Philip J. Edwards, Andrew P. King, Calvin R. Maurer, Jr., Darryl A. de Cunha, David J. Hawkes, Derek L. G. Hill, Ron P. Gaston, Michael R. Fenlon, A. Jusczyzck, Anthony J. Strong, Christopher L. Chandler, and Michael J. Gleeson. Design and evaluation of a system for microscope-assisted guided interventions (MAGI). *IEEE Transactions on Medical Imaging*, 19(11):1082–1093, 2000.

[Eng02] Klaus Engel and Thomas Ertl. Interactive high-quality volume rendering with flexible consumer graphics hardware. In *Eurographics STAR (state-of-the-art report)*, September 2002.

[Fic04] Gabor Fichtinger, Anton Deguet, Ken Masamune, Emese Balogh, Gregory Fischer, Herve Mathieu, Russell H. Taylor, Laura M. Fayad, and S. James Zinreich. Needle insertion in CT scanner with image overlay – cadaver studies. In *Proceedings of the 7th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 3217 of *Lecture Notes in Computer Science*, pages 795–803, September 2004.

[Fic05a] Gabor Fichtinger, Anton Deguet, Gregory Fischer, Iulian Iordachita, Emese Balogh, Ken Masamune, Russell H. Taylor, Laura M. Fayad, Michelle de Oliveira, and S. James

Zinreich. Image overlay for CT-guided needle insertions. *Computer Assisted Surgery*, 10(4):241–255, July 2005.

[Fic05b] Gabor Fichtinger, Anton Deguet, Ken Masamune, Emese Balogh, Gregory S. Fischer, Herve Mathieu, Russell H. Taylor, S. James Zinreich, and Laura M. Fayad. Image overlay guidance for needle insertion in CT scanner. *IEEE Transactions on Biomedical Engineering*, 52(8):1415–1424, August 2005.

[Fig01] Michael Figl, Wolfgang Birkfellner, Johann Hummel, Rudolf Hanel, P. Homolka, Franz Watzinger, Felix Wanschitz, Rolf Ewers, and Helmar Bergmann. Current status of the Varioscope AR, a head-mounted operating microscope for computer-aided surgery. In *Proceedings of the IEEE and ACM International Symposium on Augmented Reality (ISAR)*, pages 20–29, October 2001.

[Fig02a] Michael Figl, Wolfgang Birkfellner, Christopher Ede, Johann Hummel, Rudolf Hanel, Franz Watzinger, Felix Wanschitz, Rolf Ewers, and Helmar Bergmann. The control unit for a head mounted operating microscope used for augmented reality visualization in computer aided surgery. In *Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 69–75, September 2002.

[Fig02b] Michael Figl, Wolfgang Birkfellner, Franz Watzinger, Felix Wanschitz, Johann Hummel, Rudolf Hanel, Rolf Ewers, and Helmar Bergmann. PC-based control unit for a head mounted operating microscope for augmented reality visualization in surgical navigation. In *Proceedings of the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 44–51, September 2002.

[Fig05] Michael Figl, Christopher Ede, Johann Hummel, Felix Wanschitz, Rolf Ewers, Helmar Bergmann, and Wolfgang Birkfellner. A fully automated calibration method for an optical see-through head-mounted operating microscope with variable zoom and focus. *IEEE Transactions on Medical Imaging*, 24(11):1492–1499, November 2005.

[Fis06] Gregory S. Fischer, Anton Deguet, Daniel Schlattman, Russell H. Taylor, Laura Fayad, S. James Zinreich, and Gabor Fichtinger. MRI image overlay: Applications to arthrography needle insertion. In *Proceedings of the 14th Annual Medicine Meets Virtual Reality Conference (MMVR)*, pages 150–155, January 2006.

[Fis07] Gregory S. Fischer, Anton Deguet, Csaba Csoma, Russell H. Taylor, Laura Fayad, John A. Carrino, S. James Zinreich, and Gabor Fichtinger. MRI image overlay: Application to arthrography needle insertion. *Computer Assisted Surgery*, 12(1):2–14, January 2007.

[Fol97] J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes. *Computer Graphics Principles and Practice, Second Edition in C*. Adison-Wesley, Reading, MA, 1997.

[Fri89] E. M. Friets, J. W. Strohbehn, J. F. Hatch, and D. W. Roberts. A frameless stereotaxic operating microscope for neurosurgery. *IEEE Transactions on Biomedical Engineering*, 36(6):608–617, June 1989.

[Fuc96] Henry Fuchs, Andrei State, Etta D. Pisano, William F. Garrett, Gentaro Hirota, Mark Livingston, Mary C. Whitton, and Stephen M. Pizer. Towards performing ultrasound-guided needle biopsies from within a head-mounted display. In *Proceedings of the 4th International Conference on Vizualization in Biomedical Computing (VBC)*, pages 591–600, September 1996.

[Fuc98] Henry Fuchs, Mark A. Livingston, Ramesh Raskar, D'nardo Colucci, Kurtis Keller, Andrei State, Jessica R. Crawford, Paul Rademacher, Samuel H. Drake, and Anthony A. Meyer. Augmented reality visualization for laparoscopic surgery. In *Proceedings of the 1st International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 934–943, October 1998.

[Fuc99] Henry Fuchs and Jeremy Ackerman. Displays for augmented reality: Historical remarks and future prospects. In Yuichi Ohta and Hideyuki Tamura, editors, *Mixed Reality—Merging Real and Virtual Worlds*, pages 31–40. Springer-Verlag, 1999.

[Gen00] Yakup Genc, Frank Sauer, Fabian Wenzel, Mirhan Tuceryan, and Nassir Navab. Optical see-through HMD calibration: A novel stereo method validated with a video see-through system. In *Proceedings of the IEEE and ACM International Symposium on Augmented Reality (ISAR)*, pages 165–174, October 2000.

[Glo03] Neil Glossop, Chris Wedlacke, John Moore, Terry Peters, and Zhanhe Wang. Laser projection augmented reality system for computer assisted surgery. In *Proceedings of the 6th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 2879 of *Lecture Notes in Computer Science*, pages 239–246. Springer-Verlag, November 2003.

[Goe03] G. Goebbels, K. Troche, M. Braun, A. Ivanovic, A. Grab, K. v. Lübtow, R. Sader, F. Zeilhofer, K. Albrecht, and K. Praxmarer. ARSyS-Tricorder - Entwicklung eines Augmented Reality System für die intraoperative Navigation in der MKG Chirurgie. In *Proceedings of the 2. Jahrestagung der Deutschen Gesellschaft für Computer- und Roboterassistierte Chirurgie e.V.*, November 2003.

[Gri94] W. Grimson, T. Lozano-Pérez, W. Wells, G. Ettinger, S. White, and R. Kikinis. An automatic registration method for frameless stereotaxy, image guided surgery, and enhanced reality visualization. In *Proceedings of the IEEE Computer Vision and Pattern Recognition Conference (CVPR)*, pages 430–436, June 1994.

[Gri95] W. E. L. Grimson, G. J. Ettinger, S. J. White, P. L. Gleason, T. Lozano-Pérez, W. M. Wells III, and R. Kikinis. Evaluating and validating an automated registration system for enhanced reality visualization in surgery. In *Proceedings of Computer Vision, Virtual Reality, and Robotics in Medicine '95 (CVRMed)*, pages 3–12, April 1995.

[Gri96] W. E. L. Grimson, G. J. Ettinger, S. J. White, T. Lozano-Pérez, W. M. Wells III, and R. Kikinis. An automatic registration method for frameless stereotaxy, image guided surgery, and enhanced reality visualization. *IEEE Transactions on Medical Imaging*, 15(2):129–140, April 1996.

[Gri98] E. Grimson, M. Leventon, G. Ettinger, A. Chabrerie, F. Ozlen, S. Nakajima, H. Atsumi, R. Kikinis, and P. Black. Clinical experience with a high precision image-guided neurosurgery system. In *Proceedings of the 1st International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 63–73, October 1998.

[Gri99] W. Eric L. Grimson, Ron Kikinis, Ferenc A. Jolesz, and Peter McL. Black. Image-guided surgery. *Scientific American*, 280(6):62–69, June 1999.

[Guy00] Chris Guy and Dominic ffytche. *An Introduction to The Principles of Medical Imaging*. Imperial College Press, London, 2000.

[Hal02] John W. Haller, Laurence Clarke, and Bruce Hamilton, editors. *Report of the NIH/NSF Group on Image-Guided Interventions*, NIH/NSF Workshop on Image-Guided Interventions, September 2002.

[Har91] Robert M. Haralick, Chung-nan Lee, Karsten Ottenberg, and Michael Nolle. Analysis and solutions of the three point perspective pose estimation problem. In *Proceedings* *of the IEEE Computer Vision and Pattern Recognition Conference (CVPR)*, pages 592–598, June 1991.

[Har93] R. Haralick and L. Shapiro. *Computer and Robot Vision*. Adison-Wesley, Reading, MA, 1993.

[Har00] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, UK, 2000.

[Hay05] Mitsuhiro Hayashibe, Naoki Suzuki, Asaki Hattori, Yoshito Otake, Shigeyuki Suzuki, and Norio Nakata. Data-fusion display system with volume rendering of intraoperatively scanned CT images. In *Proceedings of the 8th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 3750 of *Lecture Notes in Computer Science*, pages 559–566. Springer-Verlag, October 2005.

[Hei97] Janne Heikkilä and Olli Silvén. A four-step camera calibration procedure with implicit image correction. In *Proceedings of the IEEE Computer Vision and Pattern Recognition Conference (CVPR)*, pages 1106–1112, June 1997.

[Hei06] Sandro-Michael Heining, Philipp Stefan, Latifa Omary, Stefan Wiesner, Tobias Sielhorst, Nassir Navab, Frank Sauer, Ekkehard Euler, Wolf Mutschler, and Joerg Traub. Evaluation of an in-situ visualization system for navigated trauma surgery. *Journal of Biomechanics*, 39(Supplement 1):209, 2006.

[Hof98] W. A. Hoff. Fusion of data from head-mounted and fixed sensors. In *Proceedings of the 1st International Workshop on Augmented Reality (IWAR)*, pages 167–182, November 1998.

[Hop02] H. Hoppe, C. Kuebler, J. Raczkowsky, H. Wörn, and S. Haßfeld. A clinical prototype system for projector-based augmented reality: Calibration and projection methods. In *Proceedings of the 16th International Congress and Exhibition on Computer Assisted Radiology and Surgery (CARS)*, page 1079, June 2002.

[Hop03] Harald Hoppe, Georg Eggers, Thorsten Heurich, Jörg Raczkowsky, Rüdiger Marmulla, Heinz Wörn, Stefan Haßfeld, and José Luis Moctezuma. Projector-based visualization for intraoperative navigation: first clinical results. In *Proceedings of the 17th International Congress and Exhibition on Computer Assisted Radiology and Surgery (CARS)*, page 771, June 2003.

[Hor99] Joachim Hornegger and Carlo Tomasi. Representation issues in the ML estimation of camera motion. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 640–647, September 1999.

[Ise97] H. Iseki, Y. Masutani, M. Iwahara, T. Tanikawa, Y. Muragaki, T. Taira, T. Dohi, and K. Takakura. Volumegraph (overlaid three-dimensional image-guided navigation) — clinical application of augmented reality in neurosurgery. *Stereotactic and Functional Neurosurgery*, 68(1–4 Pt 1):18–24, June 1997.

[Jäh97] Bernd Jähne. *Digitale Bildverarbeitung*. Springer-Verlag, 4 edition, 1997.

[Kal00] Willi A. Kalender. *Computed Tomography: Fundamentals, System Technology, Image Quality, Applications*. Publicis MCD Verlag, 2000.

[Kel82] P. J. Kelly, G. J. Alker Jr, and S. Goerss. Computer-assisted stereotactic laser microsurgery for the treatment of intracranial neoplasms. *Neurosurgery*, 10(3):324–331, March 1982.

[Kha03a] Ali Khamene, Sebastian Vogt, Fred Azar, Tobias Sielhorst, Frank Sauer, and Heinrich Niemann. Local 3D reconstruction and augmented reality visualization of free-hand ultrasound for needle biopsy procedures. In *Proceedings of the 6th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 2879 of *Lecture Notes in Computer Science*, pages 344–355. Springer-Verlag, November 2003.

[Kha03b] Ali Khamene, Frank Wacker, Sebastian Vogt, Fred Azar, Michael Wendt, Frank Sauer, and Jonathan Lewin. An augmented reality system for MRI-guided needle biopsies. In *Proceedings of the 11th Annual Medicine Meets Virtual Reality Conference (MMVR)*, pages 151–157, January 2003.

[Kha04] Ali Khamene, Jan K. Warzelhan, Sebastian Vogt, Daniel Elgort, Christophe Chefd'Hotel, Jeffrey L. Duerk, Jonathan S. Lewin, Frank K. Wacker, and Frank Sauer. Characterization of internal organ motion using skin marker positions. In *Proceedings of the 7th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 3217 of *Lecture Notes in Computer Science*, pages 526–533, September 2004.

[Kin00] A. P. King, P. Edwards, C. R. Maurer Jr., D. A. de Cunha, R. P. Gaston, M. Clarkson, D. L. G. Hill, D. J. Hawkes, M. R. Fenlon, A. J. Strong, T. C. S. Cox, and M. J.

Gleeson. Stereo augmented reality in the surgical microscope. *Presence: Teleoperators and Virtual Environments*, 9(4):360–368, 2000.

[Len88] Reimar K. Lenz and Roger Y. Tsai. Techniques for calibration of the scale factor and image center for high accuracy 3-D machine vision metrology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(5):713–720, September 1988.

[Lia01] Hongen Liao, Susumu Nakajima, Makoto Iwahara, Etsuko Kobayashi, Ichiro Sakuma, Naoki Yahagi, and Takeyoshi Dohi. Intra-operative real-time 3-D information display system based on integral videography. In *Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 2208 of *Lecture Notes in Computer Science*, pages 392–400. Springer-Verlag, October 2001.

[Lia02] Hongen Liao, Nobuhiko Hata, Makoto Iwahara, Susumu Nakajima, Ichiro Sakuma, and Takeyoshi Dohi. High-resolution stereoscopic surgical display using parallel integral videography and multi-projector. In *Proceedings of the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 2489 of *Lecture Notes in Computer Science*, pages 85–92. Springer-Verlag, September 2002.

[Lia04] Hongen Liao, Nobuhiko Hata, Susumu Nakajima, Makoto Iwahara, Ichiro Sakuma, and Takeyoshi Dohi. Surgical navigation by autostereoscopic image overlay of integral videography. *IEEE Transactions on Information Technology in Biomedicine*, 8(2):114–121, June 2004.

[Loo99] Jack M. Loomis, Roberta L. Klatzky, and Reginald G. Golledge. Auditory distance perception in real, virtual, and mixed environments. In Yuichi Ohta and Hideyuki Tamura, editors, *Mixed Reality—Merging Real and Virtual Worlds*, pages 201–214. Springer-Verlag, 1999.

[Lor93] William Lorensen, Harvey Cline, Christopher Nafis, Ron Kikinis, David Altobelli, and Langham Gleason. Enhancing reality in the operating room. In *Proceedings of IEEE Visualization ’93*, pages 410–415, October 1993.

[Lor94] W. Lorensen, H. Cline, C. Nafis, R. Kikinis, D. Altobelli, L. Gleason, and F. Jolesz. Enhanced reality in the operating room. In *Proceedings of the 2nd Annual Medicine Meets Virtual Reality Conference (MMVR)*, pages 124–127, January 1994.

[Mas95] Y. Masutani, M. Iwahara, O. Samuta, Y. Nishi, M. Suzuki N. Suzuki, T. Dohi, H. Iseki, and T. Takakura. Development of integral photography-based enhanced reality visualization system for surgical support. In *Proceedings of the Second International Symposium on Computer Aided Surgery (ISCAS)*, pages 16–17, October 1995.

[Mas96] Y. Masutani, T. Dohi, Y. Nishi, M. Iwahara, H. Iseki, and K. Takakura. Volumegraph — an integral photography based enhanced reality visualization system. In *Proceedings of the 10th International Symposium on Computer Assisted Radiology (CAR)*, page 1051, June 1996.

[Mas00] Ken Masamune, Yoshitaka Masutani, Susumu Nakajima, Ichiro Sakuma, Takeyoshi Dohi, Hiroshi Iseki, and Kintomo Takakura. Three-dimensional slice image overlay system with accurate depth perception for surgery. In *Proceedings of the 3rd International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 395–402, October 2000.

[Mas02] Ken Masamune, Gabor Fichtinger, Anton Deguet, Daisuke Matsuka, and Russell Taylor. 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 (MICCAI)*, volume 2489 of *Lecture Notes in Computer Science*, pages 77–84. Springer-Verlag, September 2002.

[May92] Stephen J. Maybank and Olivier D. Faugeras. A theory of self-calibration of a moving camera. *International Journal of Computer Vision*, 8(2):123–151, August 1992.

[Mil94] Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, E77-D(12):1321–1329, 1994.

[Mil99] Paul Milgram and Herman Colquhoun Jr. A taxonomy of real and virtual display integration. In Yuichi Ohta and Hideyuki Tamura, editors, *Mixed Reality—Merging Real and Virtual Worlds*, pages 5–30. Springer-Verlag, 1999.

[Mit00] M. Mitschke, A. Bani-Hashemi, and N. Navab. Interventions under video-augmented X-ray guidance: Application to needle placement. In *Proceedings of the 3rd International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 858–868, October 2000.

[Mou01] Fabien Mourgues, Frédéric Devernay, and Ève Coste-Manière. 3D reconstruction of the operating field for image overlay in 3D-endoscopic surgery. In *Proceedings of the IEEE and ACM International Symposium on Augmented Reality (ISAR)*, pages 191–192, October 2001.

[Mou03] Fabien Mourgues, Thierry Vieville Volkmar Falk, and Ève Coste-Manière. Interactive guidance by image overlay in robot assisted coronary artery bypass. In *Proceedings of the 6th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 2878 of *Lecture Notes in Computer Science*, pages 173–181. Springer-Verlag, November 2003.

[Nak00] Susumu Nakajima, Sumihisa Orita, Ken Masamune, Ichiro Sakuma, Takeyoshi Dohi, and Kozo Nakamura. Surgical navigation system with intuitive three-dimensional display. In *Proceedings of the 3rd International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 403–411, October 2000.

[Nav99a] Nassir Navab, Ali Bani-Hashemi, and Matthias Mitschke. Merging visible and invisible: Two camera-augmented mobile C-arm (CAMC) applications. In *Proceedings of the 2nd International Workshop on Augmented Reality (IWAR)*, pages 134–141, October 1999.

[Nav99b] Nassir Navab, Benedicte Bascle, Mirko Appel, and Echeyde Cubillo. Scene augmentation via the fusion of industrial drawings and uncalibrated images with a view to marker-less calibration. In *Proceedings of the 2nd International Workshop on Augmented Reality (IWAR)*, pages 125–133, October 1999.

[Nic05a] S. A. Nicolau, X. Pennec, L. Soler, and N. Ayache. A complete augmented reality guidance system for liver punctures: First clinical evaluation. In *Proceedings of the 8th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 3749 of *Lecture Notes in Computer Science*, pages 539–547. Springer-Verlag, October 2005.

[Nic05b] Stéphane Nicolau, Alain Garcia, Xavier Pennec, Luc Soler, and Nicholas Ayache. An augmented reality system to guide radio-frequency tumor ablation. *Journal of Computer Animation and Virtual World*, 16(1):1–10, 2005.

[Nie90] Heinrich Niemann. *Pattern Analysis and Understanding*, volume 4 of *Springer Series in Information Sciences*. Springer-Verlag, Heidelberg, 1990.

[Oko77] Takanori Okoshi. *Three-dimensional Imaging Techniques*. Academic Press Inc., Burlington, MA, 1977.

[Pai05] Dinesh K. Pai. Multisensory interaction: Real and virtual. In Paolo Dario and Raja Chatila, editors, *Robotics Research—The Eleventh International Symposium*, volume 15 of *Springer Tracts in Advanced Robotics*, pages 489–498. Springer-Verlag, 2005.

[Pfi96] Cuno Pfister and Clemens Szyperski. Why objects are not enough. In *Proceedings of the First International Component Users Conference (CUC)*, Munich, Germany, July 1996.

[Qua99] Long Quan and Zhongdan Lan. Linear n-point camera pose determination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(8):774–780, August 1999.

[Rol00] J.P. Rolland and H. Fuchs. Optical versus video see-through head-mounted displays in medical visualization. *Presence: Teleoperators and Virtual Environments*, 9(3):287–309, June 2000.

[Ros01] Michael Rosenthal, Andrei State, Joohi Lee, Gentaro Hirota, Jeremy Ackerman, Kurtis Keller, Etta D. Pisano, Michael Jiroutek, Keith Muller, and Henry Fuchs. Augmented reality guidance for needle biopsies: An initial randomized, controlled trial in phantoms. In *Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 2208 of *Lecture Notes in Computer Science*, pages 240–248. Springer-Verlag, October 2001.

[Ros02] Michael Rosenthal, Andrei State, Joohi Lee, Gentaro Hirota, Jeremy Ackerman, Kurtis Keller, Etta D. Pisano, Michael Jiroutek, Keith Muller, and Henry Fuchs. Augmented reality guidance for needle biopsies: An initial randomized, controlled trial in phantoms. *Medical Image Analysis*, 6(2):313–320, September 2002.

[RS02] Christof Rezk-Salama. Volume rendering techniques for general purpose graphics hardware. Ph.D. thesis, Technische Fakultät der Universität Erlangen-Nürnberg, 2002.

[Sau00] Frank Sauer, Fabian Wenzel, Sebastian Vogt, Yiyang Tao, Yakup Genc, and Ali Bani-Hashemi. Augmented workspace: Designing an AR testbed. In *Proceedings of the IEEE and ACM International Symposium on Augmented Reality (ISAR)*, pages 47–53, October 2000.

[Sau01a] Frank Sauer, Ali Khamene, Benedicte Bascle, and Gregory J. Rubino. A head-mounted display system for augmented reality image guidance: Towards clinical evaluation for iMRI-guided neurosurgery. In *Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 707–716, October 2001.

[Sau01b] Frank Sauer, Ali Khamene, Benedicte Bascle, Lars Schimmang, Fabian Wenzel, and Sebastian Vogt. Augmented reality visualization of ultrasound images: System description, calibration, and features. In *Proceedings of the IEEE and ACM International Symposium on Augmented Reality (ISAR)*, pages 30–39, October 2001.

[Sau02a] Frank Sauer, Ali Khamene, Benedicte Bascle, and Sebastian Vogt. An augmented reality system for ultrasound guided needle biopsies. In *Proceedings of the 10th Annual Medicine Meets Virtual Reality Conference (MMVR)*, pages 455–460, January 2002.

[Sau02b] Frank Sauer, Ali Khamene, Benedicte Bascle, Sebastian Vogt, and Gregory J. Rubino. Augmented reality visualization in iMRI operating room: System description and pre-clinical testing. In *Proceedings of SPIE's Conference of Medical Imaging 2002: Visualization, Image-Guided Procedures, and Display*, volume 4681, pages 446–454, February 2002.

[Sau02c] Frank Sauer, Ali Khamene, and Sebastian Vogt. An augmented reality navigation system with a single-camera tracker: System design and needle biopsy phantom trial. In *Proceedings of the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 116–124, September 2002.

[Sau03] Frank Sauer, U. Joseph Schoepf, Ali Khamene, Sebastian Vogt, Marco Das, and Stuart G. Silverman. Augmented reality system for CT-guided interventions: System description and initial phantom trials. In *Proceedings of SPIE's Conference of Medical Imaging 2003: Visualization, Image-Guided Procedures, and Display*, volume 5029, pages 384–394, February 2003.

[Sau06] Frank Sauer, Sebastian Vogt, Ali Khamene, Sandro Heining, Ekkehard Euler, Marc Schneberger, Konrad Zuerl, and Wolf Mutschler. Augmented reality visualization for thoracoscopic spine surgery. In *Proceedings of SPIE's Conference of Medical Imaging 2006: Visualization, Image-Guided Procedures, and Display*, volume 6141, pages 31–39, March 2006.

[Sau08] Frank Sauer, Sebastian Vogt, and Ali Khamene. Augmented Reality. In Terry Peters and Kevin Cleary, editors, *Image-Guided Interventions: Technology and Applications*. Springer-Verlag, 2008.

[Sch98] D. Schmalstieg, A. Fuhrmann, Z. Szalavari, and M. Gervautz. 'Studierstube' — An environment for collaboration in augmented reality. *Virtual Reality — Systems, Development and Application*, 3(1):37–49, 1998.

[Sch02] U. Joseph Schoepf, Frank Sauer, Ali Khamene, Sebastian Vogt, Marco Das, and Stuart G. Silverman. Augmented reality visualization for CT-guided interventions: Feasibility and phantom evaluation. In *Radiological Society of North America's 88th Scientific Assembly and Annual Meeting (RSNA)*, December 2002. Paper number: 962 in session: K05.

[Sha98] Ramin Shahidi, Bai Wang, Marc Epitaux, Robert Grzeszczuk, and John Adler. Volumetric image guidance via a stereotactic endoscope. In *Proceedings of the 1st International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 241–252, October 1998.

[Sie04a] Tobias Sielhorst, Tobias Obst, Rainer Burgkart, Robert Riener, and Nassir Navab. An augmented reality delivery simulator for medical training. In *Proceedings of the International Workshop on Augmented Environments for Medical Imaging (AMI-ARCS 2004) - MICCAI Satellite Workshop*, pages 11–20, September 2004.

[Sie04b] Tobias Sielhorst, Jörg Traub, and Nassir Navab. The AR apprenticeship: Replication and omnidirectional viewing of subtle movements. In *Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 290–291, November 2004.

[Sie06] Tobias Sielhorst, Christoph Bichlmeier, Sandro Michael Heining, and Nassir Navab. Depth perception – a major issue in medical AR: Evaluation study by twenty surgeons. In *Proceedings of the 9th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 4190 of *Lecture Notes in Computer Science*, pages 364–372. Springer-Verlag, September 2006.

[Sil99] Stuart G. Silverman, Kemal Tuncali, Douglass F. Adams, Richard D. Nawfel, Kelly H. Zou, and Philip F. Judy. CT fluoroscopy-guided abdominal interventions: Techniques, results, and radiation exposure. *Radiology*, 212(3):673–681, September 1999.

[Sta94] Andrei State, David T. Chen, Chris Tector, Andrew Brandt, Hong Chen, Ryutarou Ohbuchi, Mike Bajura, and Henry Fuchs. Case study: Observing a volume rendered fetus within a pregnant patient. In *Proceedings of IEEE Visualization '94*, pages 364–368, October 1994.

[Sta96] Andrei State, Mark A. Livingston, Gentaro Hirota, William F. Garrett, Mary C. Whitton, Henry Fuchs, and Etta D. Pisano. Technologies for augmented-reality systems: realizing ultrasound-guided needle biopsies. In *SIGGRAPH 96 Computer Graphics Proceedings, Annual Conference Series*, pages 439–446, August 1996.

[Sta03] Andrei State, Kurtis Keller, Michael Rosenthal, Hua Yang, Jeremy Ackerman, and Henry Fuchs. Stereo imagery from the UNC augmented reality system for breast biopsy guidance. In *Proceedings of the 11th Annual Medicine Meets Virtual Reality Conference (MMVR)*, pages 325–328, January 2003.

[Ste00] George D. Stetten, Vikram S. Chib, and Robert J. Tamburo. Tomographic reflection to merge ultrasound images with direct vision. In *Proceedings of the Applied Imagery Pattern Recognition Annual Workshop (AIPR)*, pages 200–205, October 2000.

[Ste01a] George Stetten and Vikram Chib. Magnified real-time tomographic reflection. In *Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 2208 of *Lecture Notes in Computer Science*, pages 683–690. Springer-Verlag, October 2001.

[Ste01b] George Stetten and Vikram Chib. Overlaying ultrasound images on direct vision. *Journal of Ultrasound in Medicine*, 20(3):235–240, 2001.

[Ste01c] George Stetten, Vikram Chib, Daniel Hildebrand, and Jeannette Bursee. Real time tomographic reflection: Phantoms for calibration and biopsy. In *Proceedings of the IEEE and ACM International Symposium on Augmented Reality (ISAR)*, pages 11–19, October 2001.

[Ste03] George Stetten, Aaron Cois, Wilson Chang, Damion Shelton, Robert Tamburo, John Castellucci, and Olaf von Ramm. C-mode real time tomographic reflection for a matrix array ultrasound sonic flashlight. In *Proceedings of the 6th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 2879 of *Lecture Notes in Computer Science*, pages 336–343. Springer-Verlag, November 2003.

[Ste05] George Stetten, Aaron Cois, Wilson Chang, Damion Shelton, Robert Tamburo, John Castellucci, and Olaf von Ramm. C-mode virtual image display for a matrix array ultrasound sonic flashlight. *Academic Radiology*, 12(5):535–543, May 2005.

[Sut65] Ivan E. Sutherland. The ultimate display. In *Proceedings of the International Federation of Information Processing Congress (IFIP)*, volume 2, pages 506–508, 1965.

[Sut68] Ivan E. Sutherland. A head-mounted three-dimensional display. In *Proceedings of the AFIPS Fall Joint Computer Conference*, volume 33, pages 757–764, 1968.

[Szy02] Clemens Szyperski, Dominik Gruntz, and Stephan Murer. *Component Software: Beyond Object Oriented Programming*. Component Software Series. Adison-Wesley, Reading, MA, second edition, 2002.

[Tra04] Jörg Traub, Marco Feuerstein, Martin Bauer, Eva U. Schirmbeck, Hesam Najafi, Robert Bauernschmitt, and Gudrun Klinker. Augmented reality for port placement and navigation in robotically assisted minimally invasive cardiovascular surgery. In *Proceedings of the 18th International Congress and Exhibition on Computer Assisted Radiology and Surgery (CARS)*, pages 735–740, June 2004.

[Tsa87] R. Y. Tsai. A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):323–344, August 1987.

[Tuc02] Mihran Tuceryan, Yakup Genc, and Nassir Navab. Single point active alignment method (SPAAM) for optical see-through HMD calibration for augmented reality. *Presence: Teleoperators and Virtual Environments*, 11(3):259–276, June 2002.

[Uch02] Shinji Uchiyama, Kazuki Takemoto, Kiyohide Satoh, Hiroyuki Yamamoto, and Hideyuki Tamura. MR platform: A basic body on which mixed reality applications are built. In *Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 246–253, September 2002.

[Vog00] Sebastian Vogt. Optische Bestimmung der Kopfposition und -orientierung zur Visualisierung eines Lichtfeldes über ein Head-Mounted-Display. Studienarbeit (Master's preparation thesis), Chair for Pattern Recognition, University of Erlangen–Nürnberg, 2000.

[Vog01] Sebastian Vogt. Anwendung von Stereoverfahren zur Verdeckungsberechnung im Bereich erweiterte Realität. Diplomarbeit (Master's thesis), Chair for Pattern Recognition, University of Erlangen–Nürnberg, 2001.

[Vog02] Sebastian Vogt, Ali Khamene, Frank Sauer, and Heinrich Niemann. Single camera tracking of marker clusters: Multiparameter cluster optimization and experimental verification. In *Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 127–136, September 2002.

[Vog03] Sebastian Vogt, Ali Khamene, Frank Sauer, Andreas Keil, and Heinrich Niemann. A high performance AR system for medical applications. In *Proceedings of the Second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 270–271, October 2003.

[Vog04a] Sebastian Vogt, Ali Khamene, Heinrich Niemann, and Frank Sauer. An AR system with intuitive user interface for manipulation and visualization of 3D medical data. In *Proceedings of the 12th Annual Medicine Meets Virtual Reality Conference (MMVR)*, pages 397–403, January 2004.

[Vog04b] Sebastian Vogt, Frank Wacker, Ali Khamene, Daniel R. Elgort, Tobias Sielhorst, Heinrich Niemann, Jeff Duerk, Jonathan Lewin, and Frank Sauer. Augmented reality system for MR-guided interventions: Phantom studies and first animal test. In *Proceedings of SPIE's Conference of Medical Imaging 2004: Visualization, Image-Guided Procedures, and Display*, volume 5367, pages 100–109, February 2004.

[Vog06] Sebastian Vogt, Ali Khamene, and Frank Sauer. Reality augmentation for medical procedures: System architecture, single camera marker tracking, and system evaluation. *International Journal of Computer Vision*, 70(2):179–190, November 2006.

[Wac06] Frank K. Wacker, Sebastian Vogt, Ali Khamene, John A. Jesberger, Sherif G. Nour, Daniel R. Elgort, Frank Sauer, Jeffrey L. Duerk, and Jonathan S. Lewin. An augmented reality system for MR image-guided needle biopsy: Initial results in a swine model. *Radiology*, 238(2):497–504, February 2006.

[Wan90] Jih-fang Wang, Vernon Chi, and Henry Fuchs. A real-time optical 3D tracker for head-mounted display systems. *ACM SIGGRAPH Computer Graphics*, 24(2):205–215, March 1990.

[Wan05] David Wang, Bing Wu, and George Stetten. Laser needle guide for the sonic flashlight. In *Proceedings of the 8th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 3749 of *Lecture Notes in Computer Science*, pages 647–653. Springer-Verlag, October 2005.

[Wei07] Eric W. Weisstein. *Eric Weisstein's World of Mathematics*. Wolfram Research, 2007. mathworld.wolfram.com.

[Wen92] Juyang Weng, Paul Cohen, and Marc Herniou. Camera calibration with distortion models and accuracy evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(10):965–980, October 1992.

[Wen03] Michael Wendt, Frank Sauer, Ali Khamene, Benedicte Bascle, Sebastian Vogt, and Frank K. Wacker. A head-mounted display system for augmented reality: Initial evaluation for interventional MRI. *RöFo — Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*, 175(3):418–421, March 2003.

[Wör01] Heinz Wörn and Harald Hoppe. Augmented reality in the operating theatre of the future. In *Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 2208 of *Lecture Notes in Computer Science*, pages 1195–1196. Springer-Verlag, October 2001.

[Zha00] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, November 2000.

Index