Projector-based Augmented Reality and Touchless Interaction to Support MRI-Guided Interventions

Dissertation

zur Erlangung des akademischen Grades

Doktoringenieur (Dr.-Ing.)

angenommen durch die Fakultät für Informatik
der Otto-von-Guericke-Universität Magdeburg

von André Mewes, M.Sc.
geb. am 16.05.1989 in Gardelegen

Gutachterinnen/ Gutachter:
Jun.-Prof. Dr. Christian Hansen
Prof. Dr.-Ing. Stefanie Speidel
Prof. Dr.-Ing. Stefan Weber

Magdeburg, den 21.11.2019

Acknowledgements

This PhD thesis presents research that has been conducted at the Research Campus *STIMULATE* at the Otto von Guericke University Magdeburg and has been funded by the German Federal Ministry of Education and Research (BMBF) under grant number 13GW0095A and by the European Regional Development Fund (ERDF) under grant number ZS/2016/04/78123.

Zusammenfassung

Bei minimal-invasiven perkutanen Eingriffen, wie z.B. der thermischen Tumorablation, bedienen sich Radiologen bildgebender Verfahren, die Auskunft über die aktuelle Position des Ablationsapplikators und die umgebenden Risikostrukturen geben. Bei MRT-gestützten Interventionen werden diese Bilddaten auf einem Bildschirm neben dem MRI dargestellt und separieren so die Informationen vom Operationsfeld, was die Hand-Augen-Koordination erschwert, die mentale Beanspruchung des Arztes erhöht und eine ohnehin schon problematische ergonomische Situation verschlimmert. Darüber hinaus ist die Software zur Steuerung des MRT und Bereitstellung der Bilddaten für die Diagnostik konzipiert, also mit vielen Funktionen, die bei Interventionen nicht benötigt werden, und kann nur mit konventionellen Eingabemodalitäten wie Trackball, Maus oder Tastatur bedient werden. Aus diesem Grund wird die Steuerung des MRT und die Interaktion mit Bilddaten oft an Assistenten delegiert, was eine Indirektion einführt, die häufig Verwirrung und Frustration verursacht und den Ablauf der Intervention stört.

In dieser Dissertation werden Lösungsansätze für diese Probleme präsentiert. Es werden das erste projektorbasierte Augmented-Reality-Nadel-Navigationssystem für den Einsatz innerhalb der MRT-Röhre zur Unterstützung MRT-geführter Interventionen sowie ein berührungsloses Gestensteuerungs-Interface zur direkten, sterilen interventionellen Interaktion mit medizinischen Bilddaten und Steuerung des MRT vorgestellt. Das Projektor-Kamera-System wird mit einem structured-Light-Ansatz kalibriert und mit dem MRT registriert, um die visuellen Informationen von zwei eigens entwickelten Nadel-Navigationskonzepten exakt mit dem Operationsfeld zu überlagern. Das berührungslose Gestenset ist metaphorisch und selbsterklärend gestaltet und wurde in zwei verschiedenen Interventionsszenarien evaluiert. Die Auswertung zeigt vielversprechende Ergebnisse in Bezug auf Genauigkeit und Gebrauchstauglichkeit. Aufgrund ihres allgemeinen Designs können die in dieser Arbeit vorgestellten Systeme und Konzepte nicht nur den Arbeitsablauf von MRT-geführten perkutanen Ablationen verbessern, sondern auch auf andere Interventionen übertragen werden. In zukünftigen Arbeiten sollten die Projektions- und Navigationsinformationen an sich durch Atmung bewegende innere und äußere

Strukturen angepasst werden, um schließlich die klinische Anwendbarkeit erreichen zu können.

Abstract

During minimally-invasive percutaneous interventions, such as thermal ablations of tumours, the radiologist relies on imaging modalities that provide information about the current pose of the ablation applicator and the surrounding risk structures. In the case of MRI-guided interventions this image data is presented on a screen next to the MRI thus separating the information from the operating field complicating the hand-eye coordination, increasing the mental demand of the physician and worsening an already challenging ergonomic situation. Furthermore, the software to control the MRI and providing the image data is designed for diagnostic scenarios, hence with many features not needed during interventions, and can only be operated with conventional input modalities such as a trackball, mouse, or a keyboard. This is why the control of the MRI and interaction with image data is often delegated to assistants, which introduces an indirection that frequently causes confusion and frustration and disturbs the workflow of intervention.

This dissertation addresses these issues by presenting the first projector-based augmented reality navigation system to be used inside the MRI bore to support MRI-guided interventions and by introducing a touchless gesture-controlled interface to interact with medical images and control the MRI directly and sterilely during interventions. The projector-camera system is calibrated with a structured-light approach and registered with the MRI to align the visual information of two developed needle navigation concepts accurately with the operating field. The touchless gesture set is designed to be metaphoric and self-describing and was evaluated in two different image-guided intervention scenarios. The evaluation shows promising results regarding accuracy and usability. Due to their general design the systems and concepts presented in this thesis may not only improve the workflow of MRI-guided percutaneous ablations, but may also be transferred to other interventional scenarios. In future works, the projection and navigation information should be adapted to moving inner and outer structures caused by breathing to eventually be able to reach clinical applicability.

Contents

	Page
1 Introduction	**1**
2 Clinical Background	**5**
2.1 Liver Tumours	5
2.1.1 Treatment of Liver Tumours	6
2.1.2 Local Ablation of Liver Tumours	7
2.2 MRI-Guided Percutaneous Thermal Ablation of Liver Tumours	10
2.2.1 General Workflow	10
2.2.2 Applicator Navigation	12
3 Technical Foundations	**17**
3.1 Augmented Reality	17
3.2 Augmented Reality Displays	19
3.2.1 Optical See-Through Displays	20
3.2.2 Retinal Displays	21
3.2.3 Video See-Through Displays	21
3.2.4 Projector-Based Augmented Reality	24
3.3 Calibration	25
3.3.1 Camera Calibration	25
3.3.2 Projector Calibration	29
3.4 Tracking Methods	36
3.4.1 Optical Tracking	37
3.4.2 Electromagnetic Tracking	42
3.4.3 Other Tracking Methods	42
3.5 Visualisation Techniques	42
3.5.1 Augmented Reality Visualisation	43
3.5.2 Depth Perception in Augmented Reality	44
3.6 Interventional Touchless Gesture Interaction	50
3.6.1 Natural User Interfaces	52
3.6.2 Gestures	52
3.6.3 Devices for Touchless Gesture Recognition	53

3.6.4 Challenges in Gesture Design 57

4 Projector-Based Augmented Reality to Support MRI-Guided Interventions 59
4.1 Related Work 60
4.2 Materials and Methods 66
4.2.1 Hardware Setup 66
4.2.2 Calibration Process 68
4.2.3 Surface Reconstruction and Visualisation 72
4.3 Needle Guidance 74
4.3.1 Navigation by Explicit Aids (*2D*) 75
4.3.2 Navigation by See-Through Vision (*3D*) 76
4.4 Evaluation 79
4.4.1 Apparatus 79
4.4.2 Study Procedure 79
4.4.3 Measures 80
4.5 Results 80
4.6 Discussion 83
4.7 Conclusion 87

5 Interventional Touchless Gesture Interaction 91
5.1 Gesture-Controlled Interactive Radiation Shield 92
5.1.1 Related Work on Interventional Touchless Interaction 92
5.1.2 System Setup 94
5.1.3 Gesture Set 97
5.1.4 Evaluation 98
5.1.5 Results 101
5.1.6 Discussion 102
5.1.7 Follow-Up 105
5.2 Touchless Gesture Control of an MRI Scanner 106
5.2.1 Related Work on MRI Gesture Control 108
5.2.2 Requirements and Setup 109
5.2.3 Evaluation 112
5.2.4 Results 116
5.2.5 Discussion 119
5.3 Conclusion 120

6 Conclusion 123

Bibliography 127

Publication List 153

Appendix Questionnaire to Evaluate the Gesture Set (Chapter 5) 163

1 Introduction

The way physicians treat tumours has come a long way in the past. What began with phytotherapy and poultice [1] to cause tumour regression through inflammation is nowadays a technically sophisticated process. In recent years, there has been a significant trend away from open surgery towards minimally-invasive interventions for certain treatments. This includes applications for endovascular procedures, e.g. endovascular aneurysm repair, as well as applications in the abdomen, especially ablation, biopsy, and even resection. Reasons for the increasing popularity of minimally-invasive interventions, especially percutaneous tumour ablations, are shorter recovery rates and thus shorter hospital stay – which enables repeated treatments – while showing comparable mortality and recurrence rates as open surgery [2, 3]. However, by far, not all open surgery tumour treatments can be substituted with percutaneous ablation. The outcomes of percutaneous ablations depend on conditions like number, size, and location of the lesions and the completeness of ablation [4, 5]. A successful percutaneous ablation therefore needs adequate, high-resolution image acquisition to support the physicians.

So far, ultrasound (US) [6–9] and computed tomography (CT) [10–13] have become established interventional imaging modalities. US is compact and inexpensive, and thus widely available [14]. Yet, US waves are reflected by bones and air, complicating the detection of underlying structures [15]. Additionally, the depth of penetration into the patient's body is relatively low, further narrowing the field of view. Furthermore, the resolution of US images is low and the dependence of its usage and movement to the imaging quality is significant [16].

CT, in contrast, provides a full view of the whole operating area. Yet, there are several downsides: it has only limited soft tissue contrast during fluoroscopy restricting the imaging of organs and it uses ionising radiation potentially harming the patient and physician [17]. The former makes it difficult to distinguish between healthy tissue and targeted lesions or necrotic tissue, respectively [18]. Thus, live image control of the applicator position is limited [15, 19]. Because the CT can acquire 2D images only on three axes, the applicator placement on angulated paths cannot be supported optimally. With the use of magnetic resonance imaging (MRI), most of the disadvantages of the CT can be overcome. Images can be acquired with better soft-tissue contrast [13, 15], in arbitrary orientations – allowing needle punctures on angulated paths –, and without emitting ionising radiation. Additionally, it is possible to determine physical parameters, such as thermometry, non-invasively. This can be used to control the progression and success of the therapy during percutaneous tumour ablations [20].

Despite its superiority in terms of imaging quality of soft tissue, the MRI has not yet succeeded in becoming the standard imaging modality for image-guided procedures due to several reasons. First, the space for the patient, instruments and physician is limited. The bore of most MRI devices measure only about 60 cm in diameter so that the patient access is difficult [21]. Effort was put into interventional open MRI scanners in a so-called double-doughnut configuration with enough space to even fit a screen inside them, or biplanar systems [22]. Unfortunately, these devices were not successful – neither in diagnostics nor interventions – due to poor image quality resulting from the low magnetic field strength caused by their structure. Second, their spatial and temporal resolution is low compared to conventional closed-bore MRI scanners used for diagnostics [21, 23, 24]. Third, not every clinic can afford an MRI – it is expensive in acquisition and its running costs are about three times higher than those of a CT. Thus, most MRI scanners are fully engaged with diagnostic scans at daytime.

Several technical improvements by the manufacturers, e.g. the enlargement of the bore up to 70 cm and better image resolution at shorter acquisition time, led to a broader variety of treatment options inside the MRI due to better patient access and faster imaging. This way, live control imaging during MRI-guided interventions allows for therapy monitoring to support a complete coagulation of tumour tissue and thus lower chance of recurrence, i.e. better patient outcome [25].

Nevertheless, despite all progress achieved with the imaging modality MRI itself, the interventional MRI (iMRI) lacks a widespread distribution, because the elaborate workflow [13] still needs better instrument *navigation*, a clear *visualisation*, and a more effective and direct *interaction* with these to increase the accuracy, facilitate the targeting of malign tissue, and speed up the intervention. The current possibilities limit the radiologist to viewing planning data and navigation information on a separate display next to the MRI device, which requires more mental effort while mapping the virtual navigation information onto the real patient [26], and leads to a less ergonomic posture. Because of the magnetic field, which restricts the use of additional in-bore tracking hardware, instrument tracking is still a research subject and not widely available, leaving the physician without clear pose information on the applicator.

Considering the interaction with the presented data and control of the imaging, the physician is required to use conventional input modalities, such as a trackball and buttons integrated into the monitor, or a foot pedal, which are often located on the other side of the patient table and may be out of reach. This is why, in practice, the interaction with the imaging device is mostly delegated to an assistant outside the scanner room, which introduces another indirection and a possible source of error and frustration [19].

It is obvious that only some of these aspects may be tackled by technical solutions directly. But ruling out the technical drawbacks may have a positive effect on the economic factors. By improving the workflow of MRI-guided interventions, treatment times can be shortened reducing costs as well as the physician's mental demand during the procedure, and possibly improving the patient outcome.

Contribution

This dissertation addresses the aforementioned workflow issues regarding navigation, visualisation, and interaction. A projector-based augmented reality (AR) instrument *navigation* approach is presented and evaluated (see Section 4), which is well applicable under the difficult conditions of the MRI, facilitates the process of ablation and biopsy needle guidance, and contributes to the improvement of the workflow of MRI-guided interventions. The presented solution enables the radiologist to treat patients under live imaging in the MRI bore and thus to benefit from all its advantages. It serves as a general purpose system enabling visual information to be shown directly in the operating field aligned with the patient. In addition, a needle navigation scenario with two *visualisation* concepts using the projected AR system are presented and evaluated with regard to mental effort and usability with the objective of showing navigation cues to allow for a safe puncture during MRI-guided percutaneous tumour ablations.

Furthermore, a touchless gesture control concept is proposed and evaluated that enables physicians to directly control interventional software sterilely (see Chapter 5). A natural, metaphoric gesture set is implemented with unambiguity and usability as the main focus. The gestures are designed for features typically used with medical image viewers and MRI control, i.e. rotation of anatomical 3D models, slicing through a set of 2D images, windowing to change contrast and brightness of the images, and triggering functions such as starting, stopping, and switching MRI sequences.

2

Clinical Background

This chapter is intended to introduce into the workflow of MRI-guided interventions. Therefore, the disease entailing a large portion of MRI-guided interventions, liver cancer, is presented. After that, typical indications are discussed and the treatment of hepatic tumours is explained.

2.1 Liver Tumours

Liver cancer is the fifth common cancer and the second most frequent cause of cancer-related death globally [27]. It is much more common in men than in women and more likely in less developed countries than in more developed countries [28]. In 2015, an estimated 854,000 new liver cancer cases and 810,000 liver cancer related deaths occurred worldwide [27]. China alone accounted for about 50 % of the total number of cases and deaths. Approximately 90 % of all primary liver tumours occurring worldwide are hepatocellular carcinoma (HCC), which thus constitute a major global health problem [27]. 90 % of all HCC in the western world are mainly caused by hepatic cirrhosis [29, 30], of which most cases are induced by intense alcohol consumption [31]. Hepatitis B and C are other causes of HCC, as well as non-alcoholic liver infections or obesity [28, 32]. In addition to the HCC, 50 % of all patients with malignant diseases, e.g. colorectal or pancreas cancer, also develop metastases in the liver with a significant morbidity and mortality [3].

It should be noted that there is a difference in terminology between tumours and metastases. Metastases are tumours that have spread from the primary tumour to a secondary location, hence they are also tumours. Therefore, in this thesis, whenever it is written about "tumours", both primary and secondary tumours, i.e. metastases, are meant, whereas "metastases" only refer to secondary tumours.

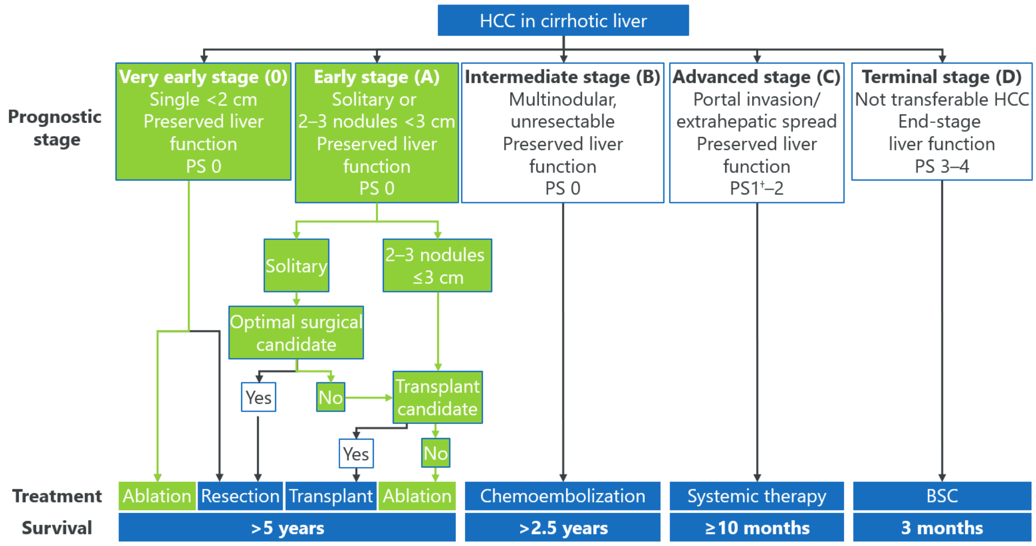


Figure 2.1: The adapted Barcelona Clinic Liver Cancer (BCLC) staging and treatment strategy according to the EASL guidelines, adapted from [5].

2.1.1 Treatment of Liver Tumours

When diagnosed with one or more tumours, a multidisciplinary tumour board discusses the best treatment options depending on the number, size, location, stage and controllability of the tumour, and the hepatic function [5] according to the guideline illustrated in Figure 2.1. This decision tree is based on the European Association for the Study of the Liver (EASL) guidelines, which focus on HCC. According to Pereira [3] and Gillams et al. [33], liver metastases and primary tumours have common criteria for the individual indications, therefore the therapy decision process can be applied to liver metastases in a similar form. The treatment options include systemic therapy, liver resection (hepatectomy), liver transplantation and ablation [34]. Systemic therapy implies side effects like hand-foot syndrome, nausea, vomiting, and worsening of liver function [34]. Resection and transplantation are radical surgical interventions and many cirrhosis patients are not suitable for this kind of therapy. A contraindication for resection is tumour-related macrovascular invasion at segmental or sub-segmental level. Transplantation shows the highest chance of cure, but is only performed on patients within the Milan criteria that have no extrahepatic metastases or vascular invasion [5].

In practice, however, hepatectomy is still the mainstay of liver tumour treatment as it shows the best outcomes in well-selected candidates. This especially applies to patients with tumours of > 2 cm in size, when the hepatic function is preserved, and sufficient remnant liver volume is maintained [5].

However, percutaneous ablation, on the other hand, is often a valuable alternative, particularly for older patients or those with a weak hepatic function or multiple small tumours [5] and even for large tumours [35]. It exhibits a lower mortality rate, lower cost, and shorter hospital stays for patients compared to surgery. When supported by imaging modalities the course and thus success of the ablation can be monitored [36].

2.1.2 Local Ablation of Liver Tumours

Thermal ablation is the local application of high or low temperatures to cause irreversible cell injury and eventually apoptosis and coagulative necrosis (cell death by denaturing structural proteins and lysosomal enzymes). Percutaneous thermal ablation can be used to treat different tumour types, e.g. liver, kidney, lung and bone cancer, but also soft tissue tumours of the breast, adrenal glands, head, and neck [36]. The increasing availability of cross-sectional imaging in the 1990s led to a rapid advance of this kind of tumour therapy. This is due to several advantages over conventional surgical resection: it offers lower morbidity, surrounding tissue is better preserved, the progress of the therapy can be visually monitored, patients, who are otherwise not suitable for surgery, can be treated, and hospital stay and costs are shortened [3]. The main goal is a *complete ablation* of all tumour cells, because otherwise the probability of local recurrence is high [34].

Commonly used thermal ablation techniques are radiofrequency ablation (RFA), microwave ablation (MWA) (both high-temperature-based), and cryoablation, which is based on low temperatures. Newer methods include high-intensity focused ultrasound (HIFU), laser ablation, and irreversible electroporation (IRE). In principle, these are similar to high-temperature-based ablation. HIFU is a non-invasive hyperthermic modality and uses multiple focused US beams to heat up cells in a selected area up to 60 °C. This way, acoustic energy causes coagulative necrosis, while the acoustic pressure leads to the collapse of cells [36, 37]. Laser ablation uses electromagnetic heating similar to RFA and MWA, but more precisely and efficiently. However, because light is scattered and absorbed, the tissue penetration depth is low, so that only small areas of 1 to 2 cm^2 can be ablated [36]. IRE primarily does not use thermal energy, but generates an electric field instead to irreversibly damage cell membranes [36, 38].

RFA, MWA, HIFU, and laser ablation cause focal hyperthermic injury to the ablated cells affecting the tumour microenvironment. The cell destruction is achieved in two phases through direct and indirect mechanisms [36]. Thermal-energy ablated tumours can be thought of as consisting of three zones: the central zone immediately beyond the applicator tip that directly undergoes coagulative necrosis, a peripheral, transitional zone of sublethal hyperthermia, that is mainly affected by thermal conduction and undergoes apoptosis or recovering from reversible injury,

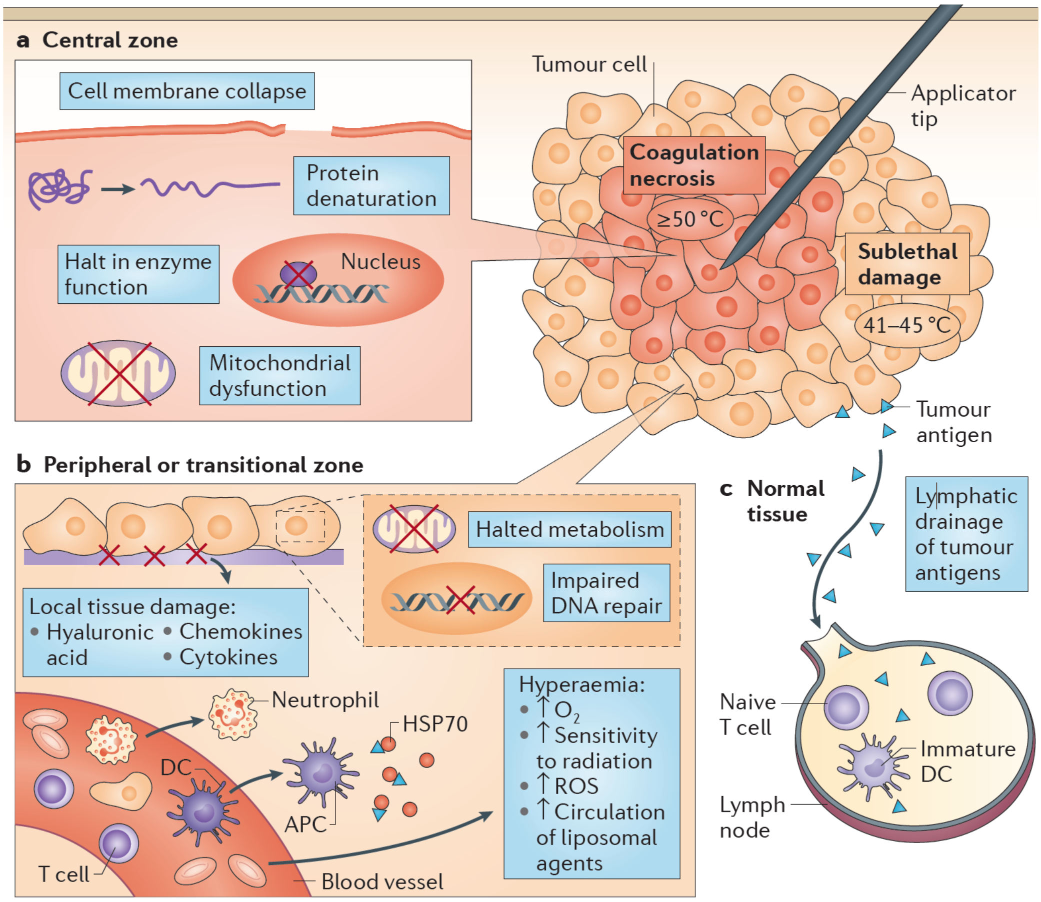


Figure 2.2: Zones of hyperthermic ablation and its effects on them [36]. ©2014 Springer Nature

and the surrounding tissue unaffected by the ablation (Figure 2.2) [36]. In the central zone, the direct damage depends on the applied thermal energy, the rate of application, and the thermal sensitivity. It happens already at temperatures of 40 to 45 °C after prolonged exposure (30 to 60 min). At temperatures above 60 °C the irreversible damage is achieved much faster by rapid protein denaturation that leads to coagulative necrosis. Secondary heat-induced damages are changes in cell membrane fluidity and permeability and thus a facilitated diffusion across the cell membrane, mitochondrial dysfunction, disturbed DNA replication through denaturation of crucial replication enzymes, and indirect, delayed cellular damage due to vascular damage or stimulation of an immune response [36, 39].

Percutaneous RFA is the insertion of one or more radiofrequency electrodes into the tumour under image guidance, e.g. with the help of US, CT or MRI. A high-frequency alternating current causes ions in the tissue to oscillate and produce frictional – or resistive – heating at temperatures of 60 to 100 °C. This way, the aforementioned hyperthermic mechanisms cause cell death by irreversible injury. Temperatures greater than 100 °C are less effective, as the tissue desiccates, which increases the tissue impedance and inhibits further electrical conduction in the remaining tissue [36, 40]. Another critical effect is the heat-sink effect occurring in proximity to larger blood vessels. The flowing blood transports heat away from the ablation zone and dissipates the hyperthermia thus decreasing the efficacy of the RFA and increasing the risk of an incomplete ablation. Tumours adjacent to vasculature are therefore less susceptible to thermal damage [36]. However, RFA triggers an immune activation in the peripheral ablation zone that confers increased tumour-free survival [36].

MWA uses the same mechanisms of hyperthermic injury as RFA, but with an antenna placed inside the tumour creating an electromagnetic field at a frequency of 900 to 2500 MHz. The field forces molecules with intrinsic dipoles (mostly water) within the tissue to consistently realign with this field. The permanent rotation of the molecules increases their kinetic energy and consequently their temperature. Because MWA does not rely on electric currents and conduction, temperatures above 100 °C do not decrease the efficacy of the ablation. It is therefore particularly suitable for tissues with higher impedance, such as lung and bone, and with high water contents like solid organs and tumours [36]. MWA heats up tissue faster than RFA, which makes it less prone to the heat-sink effect. Additionally, with MWA, tissue can be heated in a distance of up to 2 cm. RFA instead can only directly affect tissue a few millimetres away while the remaining tissue can only be reached by thermal conduction. The use of multiple antennas amplifies the heating effect, so that larger tumours can be ablated. By phasing the electromagnetic fields of multiple antennas constructively, the heating effect is further increased. This is not possible with RFA. However, MWA antennas are prone to overheating and thus need a cooling mechanism [36]. Another effect of this type of ablation is weak stimulation of local inflammation and thus a minimal innate and acquired antitumor immunity compared to RFA. However, the clinical outcome is still significantly better than after RFA treatment regarding overall survival rate and risk of local recurrence [36, 41].

In contrast to the aforementioned hyperthermic techniques, cryoablation uses cold injury to destroy cells. It is used for tumours of the retina, skin, prostate, kidney, liver, breast, lung and bone. A liquefied gas, such as argon, is delivered to the tumour through a trocar-type probe and creates a heat sink reducing the temperature at the distal end of the probe to approximately −160 °C when expanding and evaporating [36, 42]. A temperature of −40 to −20 °C must persist to 1 cm beyond the tumour to ensure a complete ablation [36, 43]. Similar to RFA and MWA, cryoablation affects cells in different zones and injury categories: direct cell injury, vascular injury and ischaemia, apoptosis, and immunomodulation. Apart from the obvious damages to cells and vascularity, cryoablation purportedly stimulates immunological targeting of tumour cells. Similar to RFA, there is the hypothesis that the destruction of malign tissue leaves intact tumour-specific antigens in situ stimulating an immune response against reversibly injured or even untreated tissue, so that metastatic tumours sometimes further regress after cryoablation, but at a much higher rate than after RFA or MWA [36, 44].

2.2 MRI-Guided Percutaneous Thermal Ablation of Liver Tumours

As already introduced, the MRI offers decisive advantages over other imaging modalities that play a role in percutaneous ablations of liver tumours. This is why MRI-guided percutaneous ablations are embedded in the EASL guidelines [5] as a potential curative approach for specific indications, especially for patients not suitable for surgery [25]. These include foremost very early and early-stage HCCs according to the BCLC staging system (see Figure 2.1) [45–48].

An image guidance modality needs to meet several requirements for a safe and effective ablation procedure. These include reliable visualisation of the targeted tumour and the applicator, accurate delineation of critical anatomical structures along the applicator's trajectory and adjacent to the tumour target, and the possibility of real-time imaging and multiplanar capabilities [49, 50].

Percutaneous ablations are most commonly performed under US or CT guidance, because these are widely available and relatively inexpensive [51, 52], but both do not meet the requirements mentioned earlier [53, 54]. Both have a relatively poor soft tissue contrast, in particular when imaging without contrast agent [55, 56], as is done mostly during interventions. This makes it difficult to reliably visualise the target tumour and place the applicator accurately and safely, especially for smaller tumours [25]. Even without contrast agent MRI provides a much better soft tissue contrast as US or CT. It provides real-time imaging and multiplanar capabilities in arbitrary plane orientations [49, 50, 57]. This makes it possible to target small tumours that may otherwise be difficult to reach [25]. The following sections provide an insight into the procedure of MRI-guided percutaneous thermal ablations.

2.2.1 General Workflow

To achieve the goal of a complete percutaneous ablation of a tumour with MRI guidance, multiple steps are necessary (see Figure 2.4). These steps were described in the test plan for the "Clinical Evaluation of MR-guided Microwave Ablation and

(a) The patient is prepared for the intervention. A flex coil is packed sterilely and placed on the intubated patient before performing an MRI planning scan.

(b) With real-time imaging the needle is guided to the tumour.

Figure 2.3: Workflow of an MRI-guided intervention.

Thermometry of Primary and Secondary Liver Malignancies" [58] and approved by the ethics committee, but they are also covered in other literature [13, 15, 19, 24].

First, the patient needs to be *prepared*. This includes patient education as well as patient positioning on the MRI table and an intubation anaesthesia. Then a flex coil is placed on the operating field and the patient is translated into the MRI. Second, during the *planning step*, directly before the intervention morphologic T1- and T2-weighted datasets are acquired. From the workstation in the MRI control room the anatomical data is used to plan the applicator path, i.e. identifying an optimal entry point and target position, incorporating pre-interventionally acquired image data. An optimal path is characterised by easy access, absence of risk structures and other organs, and shortness to injure as little healthy tissue as possible. Accordingly, the MRI planes are adjusted to visualise the whole applicator path and immediate surroundings. The planning data and interventional imaging data are presented on the in-room display in the intervention room. Third, the planned *entry point* needs to be found on the patient (see Section 2.2.2). Then the operating area is sterilised and the surroundings covered with surgical drape (see Figure 2.3a). The access path is locally anaesthetised. Fourth, the applicator is advanced under continuous T1-weighted magnetic resonance (MR) fluoroscopy (see Figure 2.3b and Section 2.2.2). After successfully placing the needle at the *target*, the ablation step starts. Fifth, the nodule is *ablated* for a certain amount of time, typically about 8 to 10 min for MWA, depending on the tumour size and the applicator specifications. During the ablation, the progress is monitored with thermometry imaging [59, 60]. After the ablation, an additional T1-weighted dataset is acquired to check the completeness of the ablation. If it did not succeed, the applicator needs to be repositioned and the ablation process is to be repeated. Otherwise, a

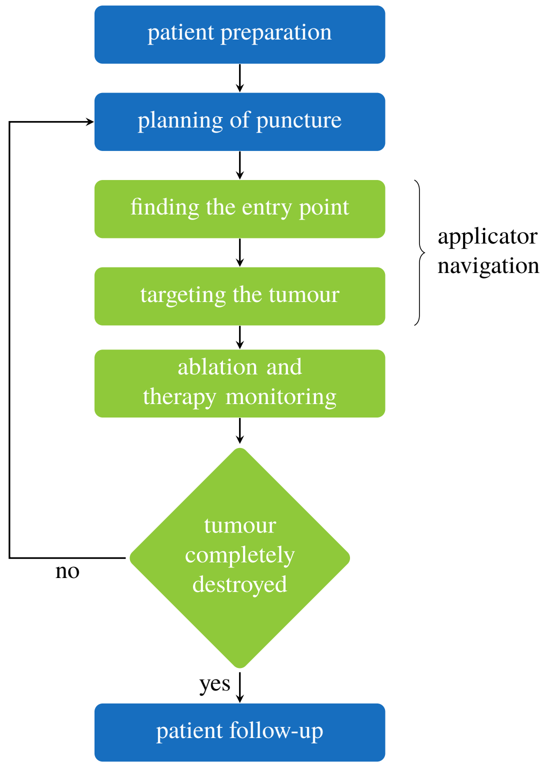


Figure 2.4: Workflow of an MRI-guided percutaneous thermal ablation. The navigation and ablation steps mark the actual intervention, the other nodes are pre- and postinterventional steps.

final control scan with contrast agent is run to verify the necrosis zone. The whole intervention time accumulates to approximately 120 to 180 min. Last, the patient is moved out of the MRI and extubated in the wake-up room. The follow-up is set after three and twelve months.

The efficacy of the procedure depends mainly on the navigation and ablation steps in Figure 2.4: the applicator navigation, i.e. finding the entry point and targeting the tumour, the duration of the ablation, and the need for repetition due to an incomplete ablation. Several methods to navigate the applicator to the tumour are described in the following.

2.2.2 Applicator Navigation

Navigating a needle-like applicator from the entry point on the patient's skin to the tumour is split into two stages: finding the entry point and advancing the applicator to the target.

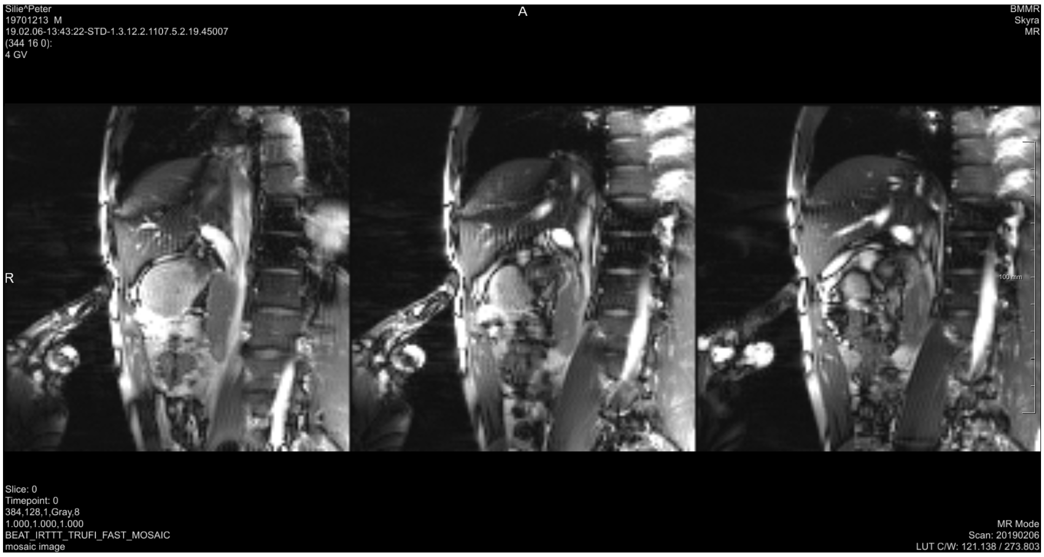


Figure 2.5: Finding the planned entry point with a finger on the patient's skin in real-time images from the MRI.

Finding the Entry Point

There are multiple methods to find the planned entry point on a patient. The most common is the finger tipping method. Under continuous imaging in the planned imaging plane the radiologist moves a finger along the patient until it can be seen in the images. Then, a finger or water-filled syringe is moved along the imaging plane to the planned entry point. For better orientation, multiple images parallel or perpendicular to the planned image plane are acquired. The finger or syringe, respectively, remains on the patient while the patient is being moved out of the MRI bore where the position is marked with a pen [19, 61–65]. A variation of this method is to carry an MRI-visible marker, e.g. nitro-glycerine capsule, to the intended entry point with the help of MR fluoroscopy and fix it with adhesive tape at this position [66]. These methods come with an inherent inaccuracy when marking the entry position or taping the capsule. Because the actual entry point may differ significantly from the planned point it is often necessary to update the acquired image planes to the actual needle path. This iterative process can soon become time-consuming and is ergonomically challenging. Reorienting the image planes also requires interaction with the MRI's software, which is cumbersome to do for the sterile radiologist (see Chapter 5). A more sophisticated process makes use of the positioning laser used for isocentre positioning at the entry of the MRI bore [13]. After planning the applicator path and setting the image planes for MR fluoroscopy, the patient is translated out of the bore for a calculated distance incorporating the distance between laser and isocentre and the distance from the isocentre to the entry point. This way, the entry point is located on the laser mark and can be measured along it with a tape measure and marked with a pen. The patient is then moved back into the bore, the imaging planes are automatically updated after readjusting the planned applicator path [13]. However, this method does not take the curvature of the body surface into account, thus leading to a position error.

Targeting the tumour

After finding the entry point on the patient, the applicator is inserted and guided along the planned path to the target position. The simplest method to achieve this is the freehand technique [13, 15, 25]. Using this technique requires continuous MRI fluoroscopy to bring the applicator to the target. For orientation, three parallel or perpendicular planes are acquired along the planned path which is to be followed by the applicator [13, 64]. The needle-like applicator is then not directly visible, but through the artefact it causes in the images. The images are presented on an in-room display not only with the original MRI's graphical user interface (GUI), but also with a special purpose interventional software that is capable of showing planning data along with the live images [13, 67]. The main advantage of the freehand method is the fast reaction on position changes of body structures with respect to the applicator. Especially the liver is much affected by breathing movement that needs to be levelled out. With MRI fluoroscopy the radiologist is even able to correct the applicator bending [50, 68]. Furthermore, the freehand navigation method does not require additional equipment, such as a tracking camera.

The workflow is comparable with US-guided percutaneous interventions [13]. Nevertheless, the manual adaption of the planned applicator path and thus the reorientation of the imaging plane is a major workflow disturbance, especially at the beginning of an intervention. This is also necessary when the applicator goes out of plane. Considering the applicator orientation, there is usually no visual guidance on the first centimetres of the puncture, because the applicator artefact is only visible inside the patient's body. Additionally, the artefact is affected by various factors, such as the applicator's material, the pulse sequence, the MRI's field strength, the applicators orientation relative to the magnetic field or receiver bandwidth, influencing the possibility to estimate the applicator position from the artefact [15, 64].

Mechanical guidance for the applicators exists as well [69] as an MRI-compatible manipulator driven by ultrasonic engines that can be controlled by the radiologist as a master-slave system directly connected to the MRI [70]. This way the imaging planes will be automatically adjusted according to the applicators pose, shortening this aforementioned time-consuming process. With mechanical guidance, it is generally possible to operate accurately, but the missing haptic feedback is a major drawback.

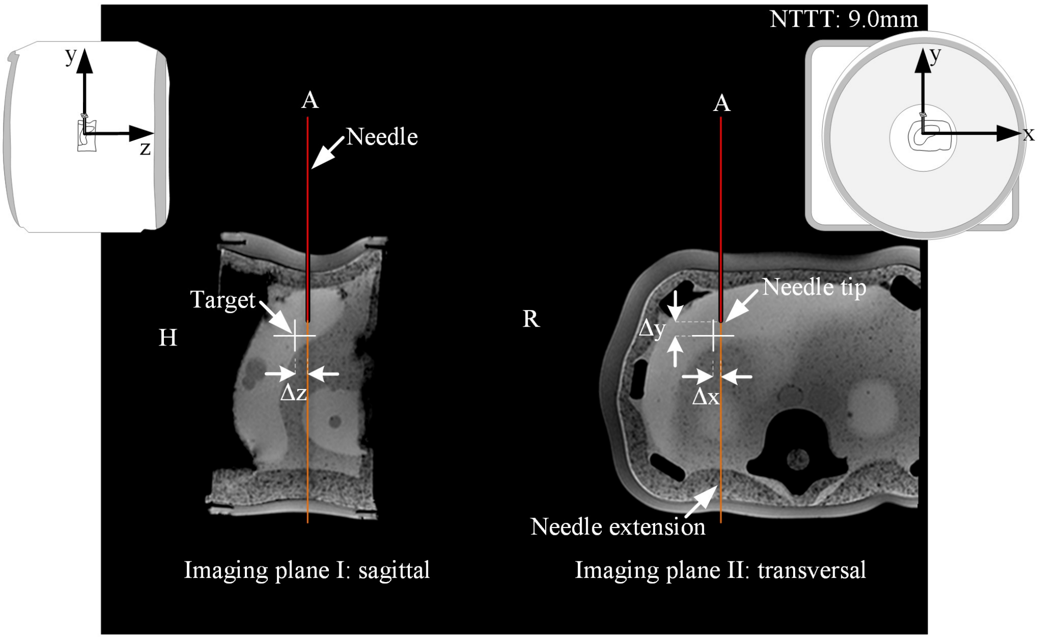


Figure 2.6: Schematic drawing of the real-time needle guidance with the Moiré phase tracking (MPT) system introduced in Kägebein et al. [71].

Automatic image alignment with the help of a needle tracking system was presented by Kägebein et al. [71] (see Figure 2.6). The tracking system is based on Moiré phase (MP) markers that are attached to the applicator tracked by a ceiling mounted camera. In a special MRI sequence the images are aligned along the applicator path. Although this method considers the need for a constant adaption of the image orientation such that the structures ahead of and around the needle are always visible, it does not support the finding of the entry point and lacks a decent field of view (FOV) of the tracking camera thus providing only a small tracking volume.

3 Technical Foundations

This chapter is intended to provide the reader with enough technical background and information on AR and instrument navigation as well as gesture-based three-dimensional user interfaces to understand the techniques presented in this thesis. Therefore, first, the term AR is defined and distinguished from similar technologies. Second, typical AR display technologies are explained. Third, the calibration methods used for calibrating projector-based AR systems are described. After that different tracking methods as well as visualisation techniques in AR and navigation systems are described. A brief introduction to natural user interfaces (NUIs) and gesture-based 3D user interfaces (UIs) completes the chapter.

3.1 Augmented Reality

AR describes a part of the mixed reality continuum, which spans between reality and virtual reality. It is also referred to as the reality-virtuality continuum (see Figure 3.1). AR is settled in between the two and describes a mix of real and virtual elements with the real objects predominating [72].

AR is not merely a combination of real and virtual objects, but rather a whole system comprised of unified real and virtual elements providing real-time interac-

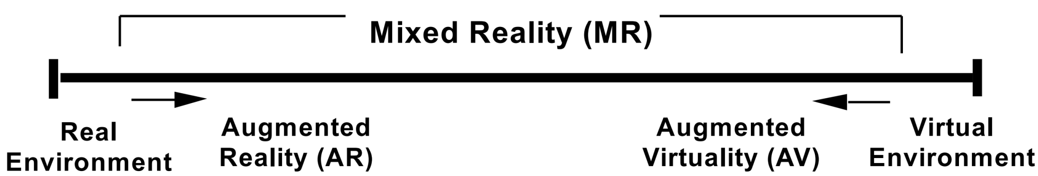


Figure 3.1: Reality-Virtuality Continuum, recreated from [72].

tivity and registering geometric objects in 3D [73]. Thus, the aim of AR is not only to overlay the reality with virtual objects, but to directly set contents into the real context by geometric registration. The contents must directly react to changes in the reality. In order to determine these changes the positions and orientations of relevant real objects must be tracked (see Section 3.4).

AR has become popular recently, although its technology has already been used in the 19th century. Back then, a technique became famous with the name "Pepper's Ghost" [74], where a large plate of glass was positioned on the stage of a theatre, such that it could reflect a person acting off-stage to create the faint illusion of a ghost next to the actors on the stage. This was an early augmentation of the real scene with a virtual character.

Contrary to the impression that AR is focused only on the visual perception, the definition of AR does not limit the type of technology used and is thus not specific to visual information. Augmentation can as well be reached with auditory or haptic enhancement of the reality. In a larger context, AR's main goal is to remove the barriers between digital interfaces and reality, thus making computer interfaces invisible and enhance the users' interaction with the real world [75].

When comparing common technologies for creating virtual reality (VR) and AR, it may be tempting to marginalise the differences between both worlds, as they may consist of the same components, such as head-mounted displays (HMDs), tracking systems, or handheld input devices. However, there are clear distinctions in the goals of AR and VR. VR separates the user inside their HMD from the real world and creates a fully immersive virtuality. AR, on the other hand, enhances the reality with digital information in a non-immersive way [76]. Consequently, a VR system depends on a wide FOV, and 3D graphics shown must be as realistic as possible to be fully immersive. Because the real world is not visible to the user, the tracking of their viewport must not be very accurate relative to the real world. AR, in contrast, depends very highly on accurate tracking for the digital information to be in the correct place, but does not require a very large FOV, because the viewport can be adapted correctly, and also works well with very simple graphics, such as arrows in navigation applications [76].

To create an AR, several components are necessary:

- Virtual contents that augment the scene,
- an AR display for the virtual contents to be displayed,
- object tracking in order to correctly register the real scene with virtual objects.

Because these depend on the use case and environment in which some techniques may be more ore less useful, the different forms of AR displays, tracking usually used for AR, but also for instrument navigation, registration methods, and visualisation aspects are explained in the following sections.

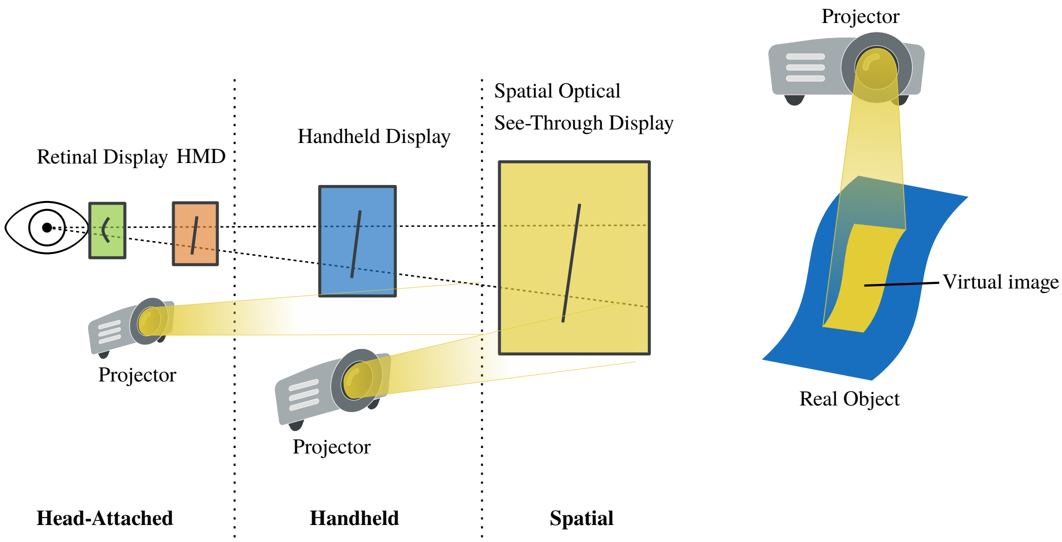


Figure 3.2: Different display categories used to create AR: head-attached, handheld, and spatial, recreated from [77].

3.2 Augmented Reality Displays

AR displays use a set of optical, electronic, and mechanical components to generate images in between the observer's eyes and a physical object to be augmented [77]. Depending on the optics used the image is formed either on a plane or a more complex non-planar surface. Figure 3.2 shows the three main categories of AR displays [77]:

Head-Attached Displays Displays attached to the head present virtual images right in front of the eyes of the user, which means they are not occluded by other objects. The most common head-attached AR display are HMDs [76] that use small displays in front of the eyes and which can be video-based or optical see-through. Additionally, two other types exist: retinal displays that project images directly onto the user's retina, and head-mounted projectors that make use of miniature projectors to throw images onto the surface of real objects [77].

Handheld Displays Handheld or body-attached AR displays are mobile, personal, and sharable [76]. They are mostly video see-through displays in the form of smartphones or tablet PCs but concepts with optical see-through displays and small projectors attached to the users shoulder [78] to control a telephone or worn in the hand [79] to support applicator placement during medical interventions also exist.

Spatial Displays Spatial displays separate the technology from the user and are usually fixed to a location and limited in mobility [77]. They are integrated into the environment and can be configured to use a beam splitter to create an optical see-through visualisation, e.g. on a desktop to create an interactive AR volume [80]. Other approaches use a projector and transparent projection film to create a holographic effect similar to optical see-through displays or use direct augmentation through projector-based AR overlays of virtual images directly on the surface of a real object [76].

3.2.1 Optical See-Through Displays

Optical see-through displays use beam splitters, such as half mirrors or prisms, to combine virtual view images with the real world view. The real world view is seen with the reflection of an image from a video display [76]. Head-up displays (HUDs) on airplane cockpits or cars are examples employing such beam splitters. Besides those splitters, transparent projection films can be used to diffuse projected light on a screen to display virtual objects while the user can view through the film at the real scene. The primary advantage of optical see-through displays is providing a direct view on the world. This way, these AR displays do not suffer from limitations in resolution, eye displacement, or time delay in the real world view, which is important in safety demanding applications, such as medical interventions. However, the calibration parameters depend on the spatial relationship between the user's viewpoint and the display's image plane. These parameters may change over time, e.g. because a wearable display slides off from the original position thus causing misalignment between real and virtual objects. Therefore, accurate 3D eye tracking relative to the display is important. Because the virtual objects are rendered based on tracking results (unless the viewpoint is static), the virtual objects are displayed temporally delayed. Correct depth occlusion is challenging because virtual objects are blended semi-transparently into the users view [76, 81]. It is further impossible to add shadows of virtual objects to the scene restricting photometric registration [81]. The real world lighting conditions further affect the perceived brightness of the scene, because optical combiners have a fixed physical transparency possibly leading to unbalanced brightness [76].

The first optical see-through HMD was introduced by Sutherland in 1968 [82]. It consisted of a head-mounted, CRT-based see-through display connected to a ceiling-mounted mechanical tracking system to track the viewport and a computer with custom graphics hardware (see Figure 3.3) [82]. The system could be used to overlay the real world scene with 3D graphics and provided a FOV of 40°, which is as much as modern HMDs use.

In recent years the research interest in optical see-through AR has increased [83] with the introduction of consumer-grade HMDs, such as the Microsoft HoloLens

Figure 3.3: Sutherland's AR system with an optical see-through HMD and mechanical tracking in 1968 [82]. Figure 2 and 3 from [82] excluded due to missing copyrights.

(Microsoft Inc, Redmond, USA, see Figure 3.4) or Meta (Meta Company, San Mateo, USA). These HMDs have been used to prototypically support navigation during medical interventions, e.g. by overlaying drilling positions during neurosurgical interventions [84].

3.2.2 Retinal Displays

Retinal displays scan modulated light directly onto the retina of the human eye with low-power lasers or special LEDs instead of using a dedicated screen in front of the eyes. This produces images that are much brighter and have higher contrast [77]. The FOV can be wider as with screen-based displays and the resolution higher. While older retinal displays only provided a fixed focus [85], a more recent approach [86] produces a pupil tracked light-field that provides depth cues depending on the user's natural focus without changing the rendered content (see Figure 3.5).

3.2.3 Video See-Through Displays

A video see-through display renders virtual images perspectively correct directly into a real scene captured with a video camera, which is often attached on the back of the display. Video see-through displays are the most widely used AR displays due to the broad availability of devices supporting it, e.g. tablet PCs and smartphones [76]. However, such displays only create the illusion of a real world

Figure 3.4: The Microsoft HoloLens optical see-through display. ©cnet.com

Figure 3.5: Schematic view of a retinal display forming laser beams to be scanned directly to the user's eyes [86]. Figure 3 from [86] excluded due to missing copyrights.

(a) Patient body augmented by live US images.

(b) HMD, camera, and US probe.

Figure 3.6: HMD with attached camera to show virtual US images aligned with the real patient [87]. Figures 4 and 5 from [87] excluded due to missing copyrights.

view, because the reality is digitised and presented on a screen. This makes it easier to control the combination of virtual and real images, because both can be shown with the same optical quality and brightness [81]. Occlusion problems that arise when an object gets in front of a virtual object can be solved by using heuristics or chroma-keying techniques, so that for example the users hand may be segmented based on the skin colour to show only the non-occluded part of the virtual object. Furthermore, depth information about the scene obtained by depth cameras can be compared with the 3D information about the virtual object [76]. The lighting conditions and colour space can be derived from the real scene and applied to the virtual scene to further increase the level of immersion of the virtual elements. The biggest drawback of video see-through displays is the indirect view on the whole scene. Although there is no temporal gap between real and virtual elements in the image, there is often a delay between the reality and the AR scene presented on the screen [76]. In safety demanding applications, such as medical interventions, the limitations in terms of resolution, delay, distortion and possible eye displacement can be critical.

Video see-through have already been used to overlay a patient with needle navigation hints. Bajura et al. proposed an AR HMD to augment a patient with live US images showing the area of interest (see Figure 3.6) [87]. Das et al. proposed a video see-through HMD to augment the view on the patient during a CT-guided percutaneous intervention with CT images, basic depth cues (see Section 3.5.2), and the part of the tracked needle which is inside the body in order to support needle guidance [88].

3.2.4 Projector-Based Augmented Reality

Projector-based AR is based on projecting virtual contents onto the surface of real objects instead of using a display. This implies that physical objects can be interactively augmented by adding information about the surface, e.g. colour and structure, and by providing additional information on the real object's surface, e.g. highlighting, symbols, explanations [81]. It is not possible to add new physical structures by projection, but to illustrate hidden structures behind the surface. The projection area can be extended by using multiple projectors.

While the user usually does not have to wear display devices, the projected AR scene is limited to a fixed location where the projector can throw images [76]. The principle of adding light to the object's surface makes the virtual elements vulnerable to external lighting and shadows from objects between the projector and the surface projected to. Providing correct occlusion may be more challenging than with other display techniques, because the users view usually differs from the projectors perspective. This makes it necessary to track or fix the viewing position and observe the projection to find occluding objects. Another major drawback of this form of AR is the perspective being only correct for one observer, as it is rendered only from one point of view. Thus, objects are located at the wrong positions for all other observers. When projecting on a non-planar surface it must be kept in mind that the focus area of projectors is most often planar. Therefore, the image may not be sharp everywhere.

To visualise virtual images perspectively correct and aligned with the surface, different geometric parameters are necessary. These include the spatial viewing position of the user, a geometric representation of the surface projected to, possibly the pose of objects lying under the surface which shall be visualised, as well as the characteristics of the projector in use. The viewing position may be fixed to a known position or needs to be tracked. The pose of objects to be aligned on the surface must be tracked as well. For an overview on suitable tracking methods see Section 3.4. It is further crucial to know the correlation between the projector pixels and world coordinates to be able to project virtual data to the desired position in space. The process of projector calibration is elucidated in Section 3.3.2. Projector-based AR is often used in arts, entertainment, or marketing (see Figure 3.7). Examples from medical interventions will be shown in Section 4.1.

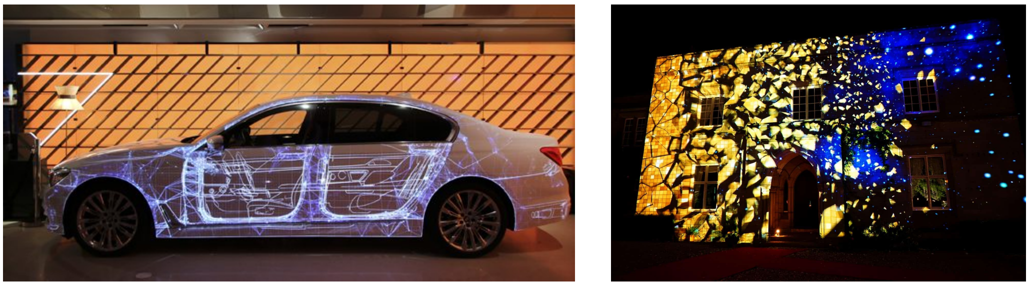


Figure 3.7: Projector-based AR. Left: in marketing, ©2019 avinteractive.com; right: in arts, ©2019 motionmapping.co.uk.

3.3 Calibration

The process of calibration may be defined as measuring inaccuracies of a device with another device of known accuracy in order to compensate these errors. Whether it is a medical instrument, a tracking camera or a projector to enhance the operating field with navigation cues: all of these devices need to be calibrated to ensure the greatest accuracy. In this chapter, camera and projector calibration methods are described.

3.3.1 Camera Calibration

Cameras are used for various purposes. In the context of AR they are often used to track optical markers or provide the user with a view of the real world that is to be augmented. Because cameras used for tracking must provide a maximally accurate image, it is necessary to calibrate the camera. Camera calibration is the process of determining internal camera geometric and optical characteristics, called *intrinsic* parameters, and the 3D pose of the camera frame to a world reference coordinate system, the *extrinsic* parameters.

Perspective Projection with the Pinhole Camera Model

The foundation of the camera calibration process is the pinhole camera model. It describes an ideal camera that assumes the camera aperture to be focused in a single point and no lenses to be used (see Figure 3.8). Light rays fall from an object in three-dimensional world coordinates through the aperture onto the image plane of the camera. There the object is mirrored. For further simplification, it is assumed that the image plane is not behind but at the same distance in front of the hole

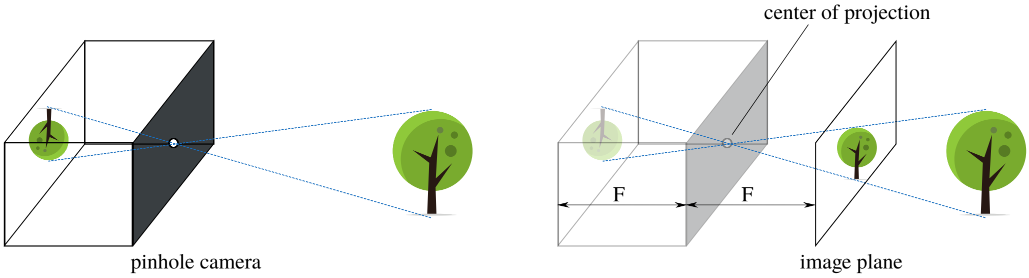


Figure 3.8: Perspective projection with the pinhole camera model. The image plane is put in front of the centre of projection to facilitate calculations.

opening. This means that the image is not rotated and the camera can be described mathematically by equations of perspective transformation [89]:

$$s\,m' = A\,[R|t]\,M' \Leftrightarrow \tag{3.1}$$

$$s\begin{bmatrix}u\\v\\1\end{bmatrix} = \begin{bmatrix}f_x & \gamma & c_x\\0 & f_y & c_y\\0 & 0 & 0\end{bmatrix}\begin{bmatrix}r_{11} & r_{12} & r_{13} & t_1\\r_{21} & r_{22} & r_{33} & t_1\\r_{31} & r_{32} & r_{33} & t_1\end{bmatrix}\begin{bmatrix}X\\Y\\Z\\1\end{bmatrix} \tag{3.2}$$

, with $f_x = Fs_x$ and $f_y = Fs_y$, F being the physical focal length, s_x, s_y being the size of the individual imager elements in pixels per millimetre, and γ being the skewness of u and v axis of the camera. This is necessary because in many cases the pixels are not square. Further be

- s – non-zero scalar
- $[u\ v\ 1]^T$ – pixel coordinates of the projected point
- A – camera matrix
- c_x, c_y– principal point: intersection of camera's optical axis and image plane
- f_x, f_y – focal length in pixels
- γ – skewness of the cameras axis u and v
- $[R|t]$ – transformation matrix from camera coordinate system to world coordinate system
- $[X\ Y\ Z\ 1]^T$ – 3D world coordinates of the projected point.

The intrinsic camera parameters are used in the camera matrix A and are independent of the scene, whereas, the extrinsic camera parameters are represented

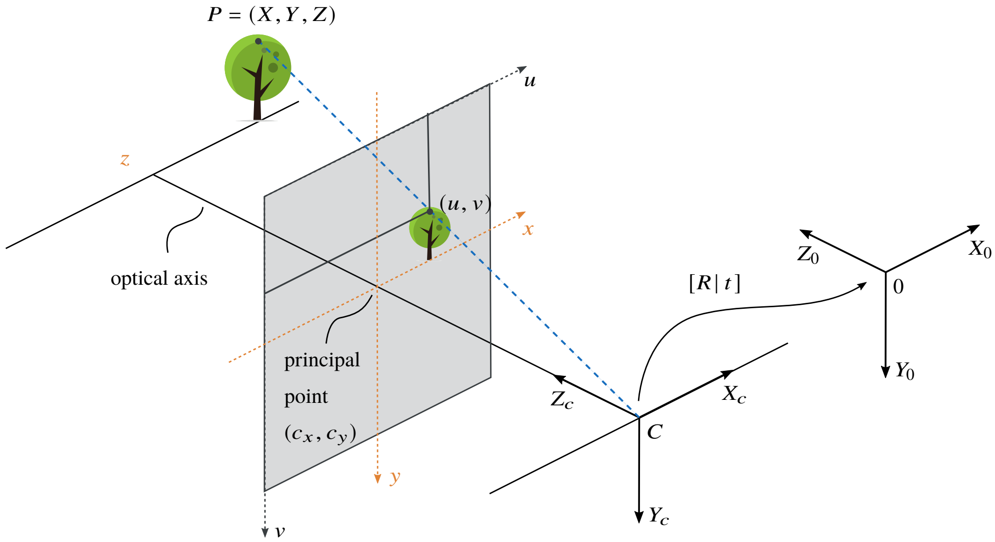


Figure 3.9: Intrinsic and extrinsic parameters of the pinhole camera.

in the transformation matrix $[R|t]$ [90]. Figure 3.9 illustrates how the different coordinate systems relate to each other.

Other than the pinhole camera, real cameras use lenses to focus more light into the camera introducing different forms of distortion. These cause the light rays to be bent and thus produce optical errors in the image. Therefore, the camera model must be extended by these distortions to map the reality more accurately. The radial and tangential distortion can be mathematically expressed by different distortion coefficients and hence be considered in the model. Because the lens distortions are independent of the camera's pose and the scene the distortion coefficients may be treated as intrinsic camera parameters, too [91]. Radial distortion is especially effective near the edges of the imager due to the shape of the non-ideal but easy-to-produce lens. It is observable as barrel or fish-eye effect. It can be corrected by [90]:

$$x_{corrected} = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \tag{3.3}$$

$$y_{corrected} = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \tag{3.4}$$

Here, x and y are the original distorted coordinates, $x_{corrected}$ and $y_{corrected}$ the new, undistorted coordinates. k_1, k_2 and k_3 are the radial distortion coefficients, r is the radius.

Tangential distortion arise from the assembly process of the camera, because the optical axis is often different from the normal of the principal point on the image sensor, thus leading to a skewed image. It can be corrected by:

$$x_{corrected} = x + [2p_1 y + p_2(r^2 + 2x^2)] \tag{3.5}$$

$$y_{corrected} = y + [p_1(r^2 + 2y^2) + 2p_2 x] \tag{3.6}$$

One of the most used approaches to find these parameters was introduced by Zhang [92]. It is based on a planar checkerboard with black and white squares and known dimensions. The algorithm can be summed up as follows [93]:

1. Images of the planar calibration pattern (checkerboard or similar) are taken in different orientations.

2. From each image the sensor points are observed, which are in direct correspondence to the points in the calibration pattern.

3. The observed points are used to estimate the homographies for each view, i.e. the linear mappings from the object points to the observed image points.

4. From the view homographies the intrinsic camera parameters are estimated in a closed-form linear solution by ignoring lens distortion. At least 3 different views of the calibration pattern are needed for a unique solution. If the sensor plane skew is assumed to be zero, 2 views are sufficient. The more images are taken, the more accurate the results will be.

5. With the intrinsic camera parameters, the extrinsic 3D parameters (rotation, translation) are calculated for each view.

6. The distortion parameters are then estimated for each view by minimising the reprojection error, i.e. the distance between the observed points and the projected points using the current intrinsic and extrinsic parameters.

7. Finally, the parameter values are used as initial guess to refine them by non-linear least-square optimisation (Levenberg-Marquardt) over all views.

For a complete and in-depth mathematical explanation please refer to Burger [93]. Other camera calibration methods were introduced by Tsai [94], Heikkila et al. [95], Weng et al. [96], and Faugeras et al. [97].

3.3.2 Projector Calibration

In order to determine the correlation between projector pixels and world coordinates the intrinsic and extrinsic projector parameters must be identified. In principle, projector calibration is similar to camera calibration: corresponding pairs of points in three-dimensional world coordinates and in projector pixel coordinates are found and then used to calculate the intrinsic parameters. The projector can be treated as an inverse camera projecting points not from world to pixel coordinates but from pixel to world coordinates.

However, the projector cannot capture a calibration pattern on its own. Therefore, an additional camera is needed for the calibration process. The projector and the camera are combined to a projector-camera system. There are a number of approaches that calibrate the camera first with the Zhang [92] method [98–103]. To this end, a calibration plane with a checkerboard attached to it is used (see Section 3.3.1). With the extrinsic parameters the pose of the board and thus the pose of the calibration plane is known, onto which then a known checkerboard is projected with the projector (see Figure 3.10). After calibrating the camera, the projected checkerboard is captured, of which the corners are projected from 2D camera coordinates to 3D world coordinates. This is possible because the calibration plane's pose is known, so that the intersections between the plane and the rays going through the checkerboard corner points in camera coordinates can be calculated. Then, these intersection points in world coordinates can be correlated with the corner points in the projector's image coordinates. Finally, these point pairs are used to calculate the intrinsic and extrinsic projector parameters similar to the process of camera calibration in Section 3.3.1 [98]. The major drawback of this approach is the dependency from the accuracy of the camera calibration. It affects the overall calibration error when world coordinates are found, which are then assigned projector correspondences. Even small camera calibration errors may result in large projector calibration reprojection errors and thus an inaccurate system calibration [104].

Another approach is to create virtual images, i.e. images from the projector's perspective, with an uncalibrated camera capturing a projected calibration pattern on a fixed plane for different projector poses [105–107]. However, moving the projector during calibration is inconvenient and in some cases not feasible.

A third projector calibration method is to iteratively adjust a projected pattern to overlap a printed calibration pattern [108–110]. An uncalibrated camera is used to capture the overlapping images. Nevertheless, both patterns must be clearly distinguishable which is why the patterns are coloured differently. Therefore, an additional colour calibration is mandatory, because otherwise printed and camera colours would unlikely match [104].

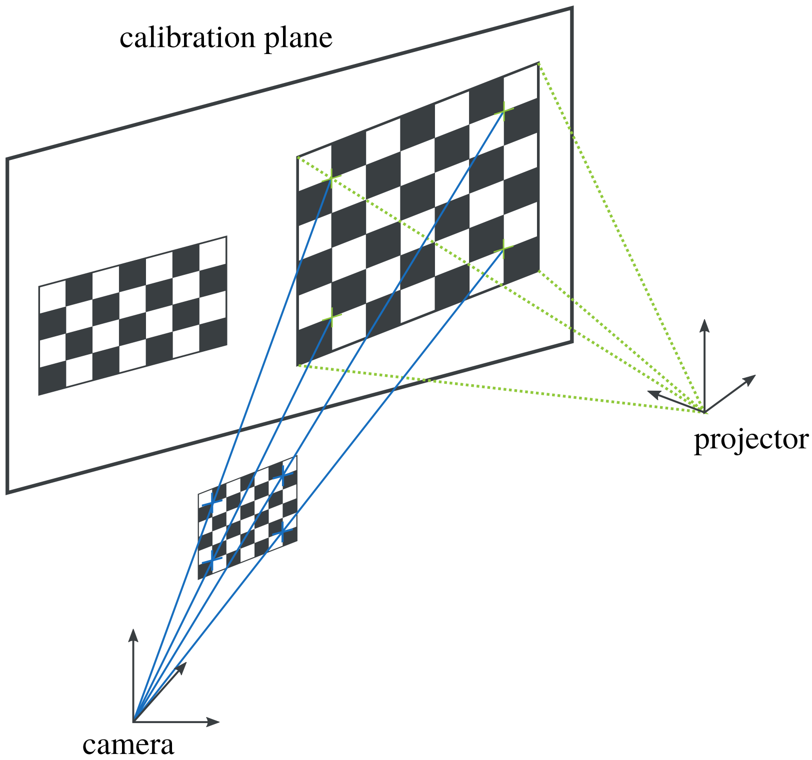


Figure 3.10: The projector is calibrated by projecting a known pattern onto a plane with known pose. With the calibrated camera the feature points of the calibration pattern are extracted and projected into 3D coordinates on the plane. With the corresponding points in world and projector coordinates the intrinsic and extrinsic parameters are calculated. Recreated from [98].

Some of these methods ([100, 105, 106, 108]) determine one global homography transformation between the calibration pattern and the projector image plane. This renders modelling of non-linear distortions impossible, because homographies are linear operators [104].

The aforementioned drawbacks are addressed in the method introduced by Moreno et al. [104]. There, structured light is projected on a planar calibration pattern observed by one camera to estimate local homographies, which are transformations from each corner point of an observed checkerboard in camera coordinates to projector pixel coordinates. The algorithm is as follows:

1. Detect the checkerboard corner locations for each plane orientation in completely illuminated images.

2. Estimate direct and global light components with high frequency gray code patterns projected onto the calibration plane.

3. Decode gray code patterns into projector row and column correspondences by means of robust pixel classification.

4. Estimate homographies for each checkerboard corner in camera coordinates to projector pixels on the basis of all correctly decoded projector pixels in the near neighbourhood of the respective corner.

5. Transform corner locations from camera coordinates to projector coordinates using the local homographies.

6. Calibrate the camera as described in Section 3.3.1.

7. Calibrate the projector with the transformed checkerboard corner points in projector pixel coordinates.

8. Estimate the stereo extrinsic parameters of camera and projector.

9. Optimise the intrinsic and extrinsic camera and projector parameters to minimise the overall reprojection error.

Some of these steps have not been discussed yet and need further explanation.

Light Component Separation and Grey Code Pattern Decoding

In Step 2, two kinds of light components that add to the total light intensity are estimated: the direct light L_d, ideally caused only by the light of one projector pixel for each camera pixel, and the indirect, global light L_g due to interreflections, i.e. light received by a surface point after reflection by other scene points, as well as volumetric scattering, subsurface scattering or light diffusion by translucent surfaces [111]. A robust separation of the light components is crucial, because errors in wrongly classified pixels propagate directly the gray code decoding step (Step 3) [104]. For decoding, a method introduced by Xu et al. [112] is used, which for its part incorporates the pixel separation method proposed by Nayar et al. [111]. These methods are also needed to reconstruct the projection surface.

Separation of the light components is done based on high frequency binary patterns, i.e. thin black and white stripes covering half of the projected image each. The method assumes only one light source being used to illuminate each surface patch. While illuminated surface patches L^+ are composed of direct light and global light, unlit surface patches L^- comprise only global light. With a fraction of activated projector pixels (α) follows [111]:

$$L^+ = L_d + \alpha L_g \tag{3.7}$$

$$L^- = (1 - \alpha) L_g \tag{3.8}$$

If a scene is lit by a high frequency pattern twice, the second image being the complementary version of the first, so that first non-illuminated patches are illuminated in the second pattern, the global light component of a patch can be computed. With knowledge about L_g the direct light component L_d can be computed.

However, this idealisation presumes that no direct projected light falls onto a surface patch intended to be unlit. In real scenarios, projectors behave differently by emitting a small amount of light even for deactivated pixels. This brightness can be described by a fraction β of the brightness of an activated pixel, with $0 < \beta < 1$. To maximise the sampling frequency and distribute the light in a scene evenly, α is set to 0.5 [111], so that

$$L^+ = L_d + (1 + \beta)\frac{L_g}{2} \tag{3.9}$$

$$L^- = \beta L_d + (1 + \beta)\frac{L_g}{2}. \tag{3.10}$$

With this separation method it is possible to decide whether a pixel is illuminated or not based on a set of rules [112] described in the following. First, the potential interval P of intensity values p is defined. For an 8-bit camera the values span from 0 to 255 . This interval is subdivided into P_{on} and P_{off}, the according intensity values for activated and deactivated pixels. To classify the pixels accurately, the lower and upper bound of these intervals must be determined. It is possible to set simple thresholds to compare the intensities of all pixels against. A more accurate approach is to adapt the threshold by averaging the intensity between a fully illuminated and an non-illuminated image for each pixel. However, these

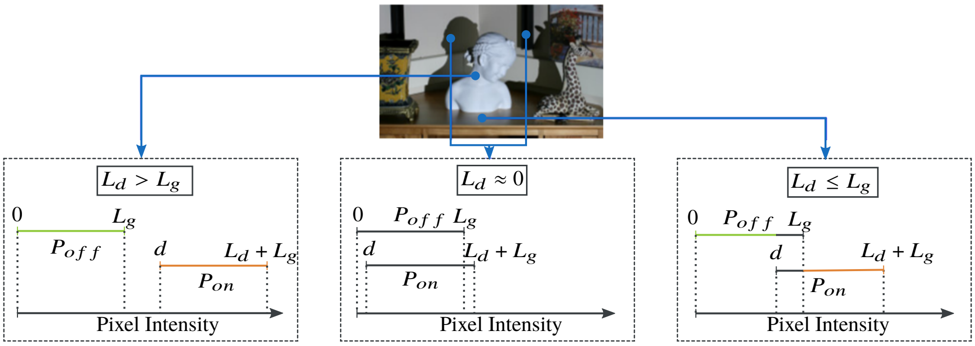


Figure 3.11: Pixel classification scenarios. a) Completely separated intervals when $L_d > L_g$. b) Indistinguishable intervals when $L_d \approx 0$. c) The intervals overlap when $L_d \leq L_g$. Recreated from [112].

methods assume that P_{on} and P_{off} do not overlap. If there is strong indirect lighting, this may not be true thus leading to incorrect classification. Furthermore, pixels never illuminated by the projector must be rejected to guarantee correct decoding.

By separating the direct and global illumination of a surface patch the pixel intensity ranges can be set to

$$\begin{aligned} P_{on} &\subseteq [L_d, L_d + L_g] \\ P_{off} &\subseteq [0, L_g] \end{aligned}$$

In the case $L_d > L_g$ (see Figure 3.11a) the two intervals are completely separated, so that the first rule can be defined as [112]:

$$\begin{aligned} p < L_g &\rightarrow \text{pixel is } \textit{off} \\ p > L_d &\rightarrow \text{pixel is } \textit{on} \\ otherwise &\rightarrow \text{pixel is } \textit{uncertain} \end{aligned}$$

If the surface point is not visible from the projector's point of view, it is not illuminated and results in L_d being close to zero (see Figure 3.11b). However, P_{on} and P_{off} cannot be distinguished anymore and thus the pixel must be discarded. This also happens when the direct light component is too small compared to the indirect light. To avoid errors, a minimum threshold m is applied to the direct light component to filter such uncertain pixels as a second rule:

$$L_d < m \rightarrow \text{pixel is } \textit{uncertain}$$

In the case $L_d \leq L_g$ (see Figure 3.11c) both intervals P_{on} and P_{off} overlap near the middle range. Hence, this interval is excluded by the third rule [112]:

$$\begin{aligned} p < L_d &\rightarrow \text{pixel is } \textit{off} \\ p > L_g &\rightarrow \text{pixel is } \textit{on} \\ otherwise &\rightarrow \text{pixel is } \textit{uncertain} \end{aligned}$$

Finally, these rules may be combined. Because a complementary pattern is projected for every gray code pattern, a consistency criterion is added: if a pixel is classified as *on* in the first pattern with a bigger direct than global component, it must not be *on* in the other; if its intensity is smaller than the indirect light component in the first pattern, its complement must be greater than the indirect component. All single rules are combined to the following dual pattern classification decision rules [112]:

$$\begin{aligned} L_d < m &\rightarrow \text{pixel is } \textit{uncertain} \\ L_d > L_g \wedge p > \bar{p} &\rightarrow \text{pixel is } \textit{on} \\ L_d > L_g \wedge p < \bar{p} &\rightarrow \text{pixel is } \textit{off} \\ p < L_d \wedge \bar{p} > L_g &\rightarrow \text{pixel is } \textit{off} \\ p > L_g \wedge \bar{p} < L_d &\rightarrow \text{pixel is } \textit{on} \\ \text{otherwise} &\rightarrow \text{pixel is } \textit{uncertain} \end{aligned}$$

With the different pixels being classified *on* or *off* in the different known projected patterns with decreasing stripe width it is possible to derive a projector row and column to each camera image coordinate and thus to assign a projector pixel to a corresponding camera pixel. This process is called decoding. Sometimes cameras have a higher resolution than the projector so that different camera pixels refer to the same projector coordinate. Then these camera pixels are grouped and the centre coordinate is taken as correspondence [104, 112].

Local Homographies

A homography is the transformation between corresponding points in two images of the same scene. In computer vision, this can be, for example, the transformation between a known printed or projected calibration pattern and an observing camera. During camera calibration it is used to describe the pose of the calibration board and the camera's optical centre. Moreno et al. [104] extend the concept of such a global homography by the introduction of local homographies – one for each of the checkerboard corners (see Figure 3.12). This means a local homography is only

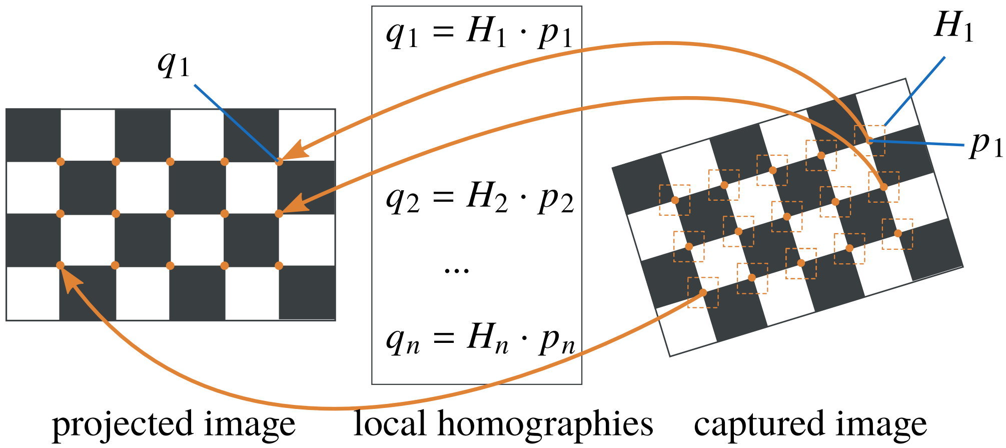


Figure 3.12: Calculation of local homographies between the projected and the captured pattern image. Recreated from [104].

valid to transform one point, in this case a chessboard corner, in camera coordinates to projector coordinates, not for any others. Therefore, local homographies enable the modelling of non-linear distortions of the projector.

To increase the robustness to small decoding errors the homographies are overdetermined through estimating them from a small neighbourhood around the corner with more points than required. All points in a patch of defined size centred at the corner location are included. Each local homography $\hat{H}$ is the result of a minimisation of the sum of squared distances between the decoded projector coordinates q and the camera pixel coordinates p of a corner point in each patch transformed by a homography H [104]:

$$\hat{H} = \arg\min_{H} \sum_{\forall p} \| \; q - Hp \; \|^2 \tag{3.11}$$

$$H \in \mathbb{R}^{3\times 3}, \; p = [u, v, 1]^T, \; q = [col, row, 1]^T$$

The corner $\bar{p}$ located at the centre of the patch is then translated to the projector coordinate $\bar{q}$ applying the local homography $\hat{H}$:

$$\bar{q} = \hat{H} \cdot \bar{p} \tag{3.12}$$

This process is repeated for all checkerboard corners. Then, with all calibration points known in projector coordinates, the intrinsic projector calibration is done analogously to the camera calibration process [104].

Stereo Calibration

So far, the intrinsic parameters for the camera and the projector are found. The world coordinates are identified with camera coordinates. Finally the projector's pose is to be found in world coordinates. This is done by stereo calibration, which yields the rotation and translation between the projector and camera image plane. This is usually done for camera-camera systems in order to be able to triangulate point positions, but it is also possible for projector-camera systems since the object points, i.e. the checkerboard corner points in world coordinates, and their representation in camera and projector points are known.

The rotation R and the translation t between the projector and the camera are calculated separately. Because P may be put into the respective camera or projector coordinates ($P_c = R_c P + t_c$ and $P_p = R_p P + t_p$) using the single-camera calibration, and the two views of the same point are related by $P_c = R^T \left(P_p - t\right)$, the rotation and translation between projector and camera can be computed with

$$R = R_p \left(R_c\right)^t \tag{3.13}$$

$$t = t_p - Rt_c. \tag{3.14}$$

This is repeated for all chessboard views yielding multiple rotations and translations which are then optimised with the Levenberg-Marquardt algorithm to minimise the reprojection error.

3.4 Tracking Methods

In order to overlay virtual information in the correct position and with the correct orientation in the real world setting, the poses of objects must be located in real-time with respect to a reference coordinate system. This process is called tracking. During medical interventions, such objects are instruments, e.g. ablation applicators or laparoscopes, the patient (operating field), or the user. The accuracy requirements depend on the application: when navigating an applicator to a target, virtual objects of an AR must be placed as accurately as possible to not endanger the patient. Depending on the use case and display technology different tracking approaches and methods are preferred. Two main approaches may be distinguished [113]:

Outside-In Tracking An external tracking device determines the poses of active or passive markers or other trackable objects.

Inside-Out Tracking A camera attached to the tracked object tracks external reference frames.

While outside-in tracking can be realised via the simple attachment of markers to the object to be tracked, inside-out tracking is only possible by installing more elaborate technology, such as cameras, and additional power supply and wiring. Different tracking methods exist for use in different environments and for different purposes in medical navigation and AR in general. These are discussed in the following.

3.4.1 Optical Tracking

Optical tracking methods are based on video-based tracking devices, which may be cameras of visible or infrared range in different setups. There exist different kinds of optical tracking that will be explained in the following sections.

Videometric Marker Tracking

Optical marker tracking became popular with the introduction of ARToolkit[1] in 1999 [76]. It is a camera-based tracking method detecting high-contrast printed markers that can be easily detected also in poor lighting conditions. The markers are mostly quadratic or circular. To be able to estimate a marker's position in real-world coordinates its form and dimensions must be known and the camera needs to be calibrated (see Section 3.3.1). Markers must usually be completely visible in the camera image. This is not the case if the marker is chosen too large for the camera-marker distance or if the marker is too small: then, the marker's features are imaged on too few camera pixels. In both cases, not enough marker features can be detected. The same applies for small camera resolution, the viewing angle being too flat, reflexions or shadows, or partly occlusion of the marker [81].

The main advantage of marker based tracking is their cheap production through printing and the easy application to objects and surroundings as well as easy integration into printed literature and advertisements. On the other hand, the most obvious disadvantage is the presence of the markers in every scene and the direct placement on the augmented objects, thus occluding them. Performance decreases significantly with an increasing number of markers in the camera image [81].

Some sorts of markers provide enough redundancies to enable further detection and localisation of the marker, even when partly occluded or with poor conditions, such as the ArUco markers [114]. These markers are created in dictionaries with determined marker size and number of bits (see Figure 3.13). A dictionary is optimised for maximum inter-marker distance and number of bit transitions to yield best distinguishable markers. The marker dictionary is to be configured to contain only as much markers as needed for the application. This way, false negative detections can be drastically reduced, because fewer markers mean a bigger inter-marker distance [115].

To detect the correct marker, the following steps are necessary [115]. The principle is similar for other AR marker detection libraries.

1. The camera captures a video or image.

2. The single image is converted to a greyscale image.

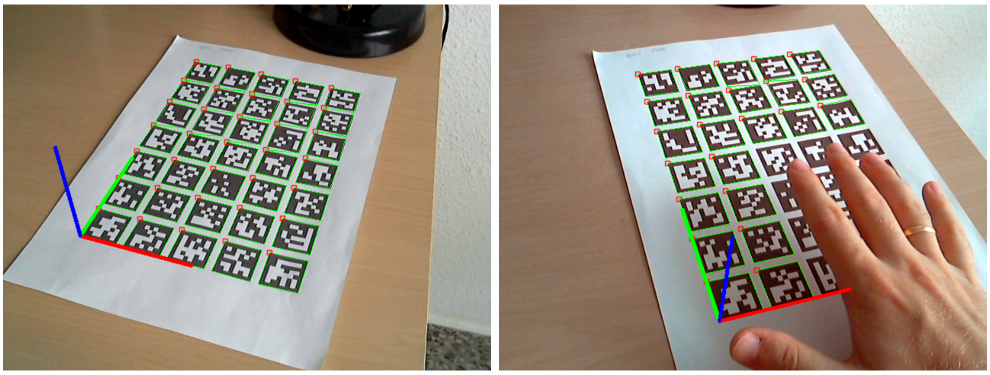
[1] https://github.com/artoolkit

Figure 3.13: Three ArUco markers in different sizes (5,6 and 8 bits) [115].

3. Local adaptive thresholding is applied to detect the most prominent contours in the image.
4. Contours are extracted and filtered.
5. With polygonal approximation rectangular contours are detected. Near contours are simplified by leaving only the outer ones.
6. To extract the inner marker code, for each marker the homography matrix is computed and the perspective projection is equalised. The marker is then thresholded, yielding a binary image which is divided into a regular grid. The border is tested for correct element values (all zeros). If accepted, the inner grid is analysed.
7. Four identifiers (inner grid) are obtained from the image, i.e. the same identifier in four orientations. These are compared with the markers in the dictionary.
8. If not found in the dictionary, error correction is applied. For each erroneous marker the distance to the markers in the dictionary is calculated. The one with the shortest distance is considered the correct one.
9. Linear regression of the marker's border pixels is used to find the corner positions. The marker pose can then be estimated by minimising the reprojection error of the corners, e.g. with the Levenberg-Marquardt algorithm.

The actual pose estimation is done by projecting the marker's corner points to the camera image points. This assumes a calibrated camera (see Section 3.3.1) and uses the intrinsic and extrinsic camera parameters. The dimensions of the marker must be known in order for this to work.

Robustness against occlusion or noise is given when using a marker board as shown in Figure 3.14. This way, more markers and thus more corners are used for



(a) ArUco marker board with all markers detected and coordinate axes.

(b) Recognition of the board although partly occluded.

Figure 3.14: Robustness of ArUco marker boards against occlusion [116].

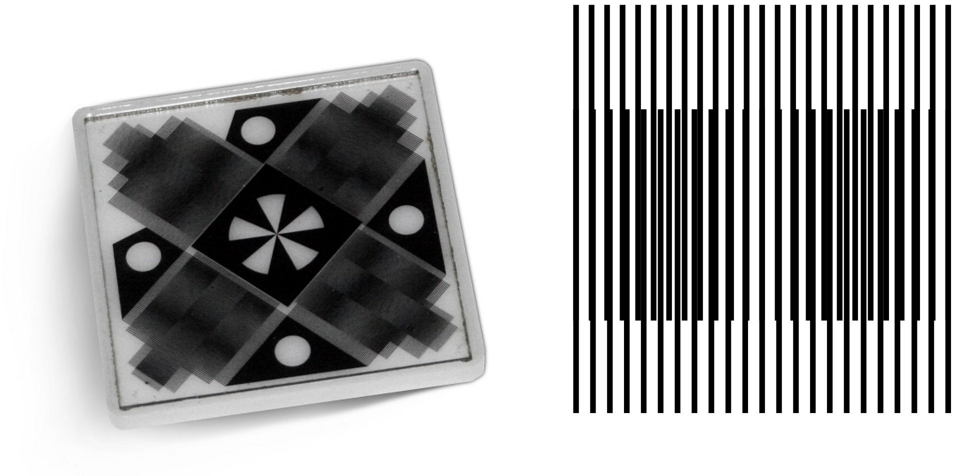
the board detection. If a few markers cannot be detected, then the board position can still be determined on the basis of the known arrangement of the individual markers still visible on the board. These properties also make ArUco boards suitable for camera calibration [115].

Moiré-Phase Tracking

A special form of optical markers are MP markers. These can be used to estimate a pose with 6 degrees of freedom (DOF) with a single camera very precisely, i.e. sub millimetre and sub degree accuracy. An MP marker consists of thee layers. It has printed planar gratings of different spatial frequencies on either side of a transparent substrate. These layers generate moiré patterns with the pattern phase depending on the marker orientation (see Figure 3.15). Similar to other optical markers, position of MP markers is tracked photogrammetrically [117]. Because of the high accuracy, these markers are especially suitable for instrument tracking in areas where workspace is too restrictive for the use of stereo cameras and retroreflective marker spheres (see Section 3.4.1).

Infrared Optical Tracking

Infrared optical tracking is the dominant tracking technology in medical interventions. It is mostly used for instrument and patient tracking when guiding applicators, but also to register virtual and real contents in AR scenes. There are two ways of infrared tracking: active and passive. When using active infrared tracking, the tracked object itself emits light from attached LEDs which is received by the cameras. For passive tracking retroreflective spherical markers are mounted onto



(a) MP marker with moiré patterns to indicate its orientation.

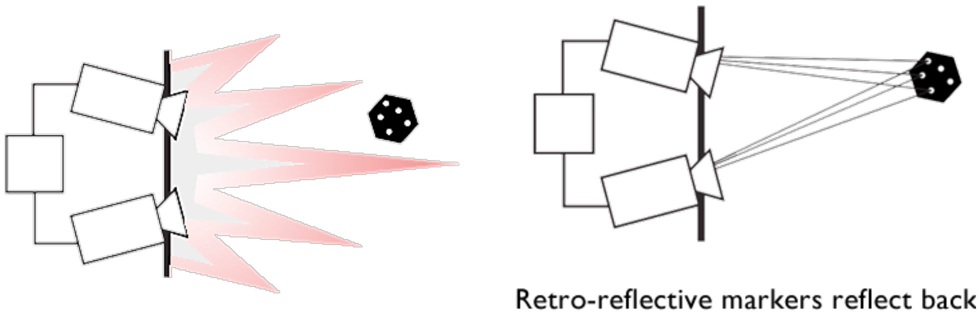
(b) Moiré pattern consisting of two layers of grates with different spatial frequencies.

Figure 3.15: Moiré marker and moiré pattern used for pose tracking. ©metriainnovation.com 2019

an object of interest. Infrared light from light sources next to the cameras is reflected from theses spheres. The positions of the active and passive markers are triangulated with the infrared cameras (see Figure 3.16). At least four markers are combined to a marker model to be able to describe a pose with 6 DOF [76]. Advantages of this method are high accuracy and fast tracking rates. It is also not influenced by conducting or metallic objects. Nevertheless, the biggest drawback is the line-of-sight problem: if a marker frame is not visible to the camera due to objects in between the tracking is interrupted. Another flaw is the relatively big size of common tracking cameras and the minimum distance for the tracking to work. Therefore, infrared tracking is not well suitable for usage in environments with limited space. Strong light from an external source may disturb the recognition of the markers. Because no electronics or cables should be mounted on the tracked object to facilitate handling, passive marker tracking is typically used in the medical context [113].

Feature-Based Tracking

Feature tracking is a tracking technique that is often used in mobile AR applications or AR supported laparoscopic interventions. To locate the pose of the camera,



(a) Light is emitted from the camera position to light up the markers.

(b) The lit markers are recognised by the cameras, triangulated and the marker frame pose is calculated.

Figure 3.16: Passive infrared light optical marker tracking [113]. ©2014 Elsevier Inc.

known features are recognised in the camera image and a transformation is estimated based on the location of these features [81].

With geometry-based tracking edges and corners are extracted from an object to serve as features and the transformation from a previous transformation is calculated from them [118]. Because of ambiguous features multiple poses may be valid. Therefore, the transform with the smallest deviation from the previous transform is chosen. The method hence relies on a good initial pose estimation, because they are incrementally calculated. Thus, this method is often combined with other, marker-based tracking techniques for a correct initialisation. This method is often used for geometric objects with too few other features [81].

Other keypoints in a camera image than edges and corners, if present, can be detected quickly and reliably. Based on a descriptor a cluster of detected keypoints can be compared to a known 2D or 3D geometry. Outliers are often removed with a RANSAC method. With the remaining feature groups the pose of the camera can be calculated. Different types of detectors exist that differ in speed and reliability, whereby independence of rotation (rotation invariance) and scaling (scale invariance) are beneficial. Otherwise, the respective keypoints must be calculated in different angles and resolutions [81]. Examples of keypoint detectors for tracking are SIFT [119] and SURF [120]. Due to the amount of detectable features these methods are very robust against occlusion. Feature-based tracking can also be used in combination with the Simultaneous Localisation and Mapping (SLAM) algorithm. When moving the camera a map is generated successively and the camera's pose is estimated at the same time [81].

3.4.2 Electromagnetic Tracking

Electromagnetic tracking is also popular in medical scenarios. Sensors inside a magnetic field of known geometry are localised. This magnetic field is created and shaped by a field generator, which implicitly encodes spatial position and orientation to the reference field. A sensor placed inside a tracked object measures the direction and strength of the magnetic flux which correlates with the distance to the source [121]. A major advantage of electromagnetic tracking is that it does not require a line of sight. Therefore, it is suitable for tracking instruments inside the human body. However, this kind of tracking suffers from different kinds of artefacts caused by instruments and other objects of conducting material or by external magnetic fields [113]. Hence, it cannot easily be used inside an MRI scanner without further adaption.

3.4.3 Other Tracking Methods

Other tracking methods exist to use with AR, which are not suitable for supporting instrument navigation during medical interventions but shall be mentioned briefly.

Global Positioning System (GPS) When tracking positions outdoors and accuracy is not important, the satellite-based GPS can be used, for example for travel navigation and information [122]. It also suffers from a line-of-sight problem to satellite, which makes it unavailable under trees or behind or inside buildings. Orientation cannot be determined.

Inertial Tracking Inertial tracking uses inertial measurement unit sensors, e.g. accelerometers, gyroscopes, and magnetometers, to measure relative pose changes and the velocity of a tracked object. It has no range limitations, line-of-sight requirements and does not interfere with remote signals. However, it is susceptible to drift [76], because the position is derived from measuring and erroneously integrating the acceleration twice. Inertial tracking is therefore often used in combination with other tracking methods by means of sensor fusion.

3.5 Visualisation Techniques

This section focuses on the special challenges posed by AR visualisation. The topic of AR visualisation in general is considered first. Afterwards, the peculiarities of depth perception are discussed, including smart-visibility techniques and illustrative visualisations. Finally, the particular challenges of projector-based AR visualisation are examined.

3.5.1 Augmented Reality Visualisation

The goal of AR is to enhance reality with virtual content. The user's view of reality is superimposed by virtual elements. These elements should fit into the real scene as well as possible. To this respect, the virtual objects must be adapted in their representation to the user's perception. They have to be displayed correctly in perspective and provided with the necessary depth information in order for the user to perceive their position correctly [123]. Further, according to Bichlmeier et al. [124], the following conditions must also be considered when visualising anatomical data:

- The view of target structures of interest must not be restricted. Objects in the line of sight must be hidden so that the target is always clearly visible.

- When using AR during interventions, instruments used must be integrated in such a way that users can perceive relative and absolute distances between instrument and anatomy.

- User interaction should be supported by visual feedback.

It is therefore important to appropriately present relevant data in order to direct the user's focus. In addition, spatial conditions must be intuitively and correctly recognised. For correct distance perception, it is important to provide virtual objects with depth information.

Kruijff et al. [125] have investigated the perception of AR visualisations and identified several problems:

Scene distortions. If virtual and real content do not fit together perfectly, it is sometimes difficult to perceive objects and their spatial dimensions correctly.

Depth distortions. If depth information is not recorded correctly, the spatial relationships between the user's first-person perspective, virtual objects, and superimposed information can no longer be interpreted correctly. This makes it difficult for users to match virtuality and reality.

Visibility. Virtual content can sometimes not be perceived correctly because it can be difficult to see due to external circumstances such as brightness or background colour.

Among these problems, Kruijff et al. determined deviations in depth perception as the most common perception problem of AR applications [125]. Therefore, the next section deals with depth perception in more detail.

3.5.2 Depth Perception in Augmented Reality

According to Kruijff et al. [125] there are several causes that may be responsible for incorrect depth perception in AR applications. For example, significant changes in the environment structure can complicate the correct registration and thus hinder the superimposition of virtual content. The colour and light conditions of the environment can also influence depth perception. Because virtual content is usually only overlaid, varying background colours and lighting change the perception of virtual objects. The colour accuracy of the visualisation is disturbed in these cases. Another cause for incorrect depth perception is masking. Due to the superimposition technology of many AR systems, such as optical see-through or projected AR, virtual objects usually appear transparent. Physical objects in the reality appear opaque behind them. As a result, virtual elements behind real objects are often perceived as lying in front of them. The user is also an error source. The human perception is based on various depth cues that result from the scene under observation, movement, accommodation, and adaptation of the eye as well as the disparity between the images of both eyes [125].

Depth cues can be differently categorised. Pictorial depth cues are part of the actual rendering, such as occlusion, image blur, size, linear perspective, texture gradient, shadows, shape from shading, height in FOV, or atmospheric perspective. Kinetic depth cues are depth cues caused by motion, such as motion parallax or accretion. There are also physiological depth cues related to the adaption of the eyes to the objects distance, e.g. accommodation, convergence, or disparity. Among these depth cues, occlusion is the one with the most striking effect on depth perception [125] and AR visualisation [123]. If virtual objects are placed in front of real objects, but are actually behind them, the spatial relationships cannot be perceived correctly. Even if other depth cues, such as binocular vision, indicate the correct position of the objects, but the order of the objects does not match, the brain cannot correctly detect the depth of the objects [123].

To achieve correct depth perception it is crucial to spatially capture the entire scene in order to be able to virtually represent the reality as well [113]. This allows for the determination of occlusions between the reality and the virtual content and therefore to render only the virtual parts lying in front of the real objects [126]. Virtual objects being situated behind real objects may also be visualised transparently as a wireframe to occlude as little real scene as possible [123].

In order to clearly visualise AR scenes, the following guidelines should be followed [127]:

Distance conveyance. Information about distances and absolute positions should be clearly visualised.

Proper motion physics. Objects should follow the laws of physics during motion to enable motion parallax.

Eliminate unneeded AR motion. Due to the motion sensitivity of perception, virtual objects should move as little as possible.

Selective or multiple cues. The accuracy of different depth cues depends on the distance of the objects to the user. Depending on the application, the depth information used should be adjusted.

Define rule space. AR visualisations should follow fixed rules. Certain objects should only appear at fixed positions and colouring should be consistent.

Careful selection of virtual content. Only a few relevant elements should be visualised in a sparse way in order to hide as few contents of reality as possible [123].

To compensate for missing depth information, it is possible to draw the region of interest as a wireframe and set other virtual objects into that context [128]. In Figure 3.17, the video feed of a laparoscopic liver resection is shown augmented by the liver as a wireframe and opaque blood vessels. The virtual objects are dynamically registered and deform in accordance to the liver. The wireframe adds context to the opaque vessels and diminishes the floating effect.

The opaque virtual objects may additionally be rendered semi-transparent through adaptive alpha-blending that depends on the distance to a reference depth [129]. This is shown in Figure 3.18. This improves the depth perception significantly. Depth perception can be enhanced by colours as well. Pseudo-chromadepth is a visualisation technique which applies a colour gradient from red to blue according to the depth of the respective structure. It was introduced by Ropinski et al. [130] and was confirmed to be suitable to render complex structures, such as vessels, by Kersten-Oertel et al. [131]. It is based on the effect that light of different wavelengths is refracted at different angles in the lens of the eye, so that this colour gradient can be used to create an illusion of depth in a flat image.

Another visualisation technique to increase depth perception is the virtual window [132]. Here, a hole in the surface of the object to be augmented is defined and creates the impression of a window through which the user can look onto virtual objects. Objects hidden by the surface are not rendered.

Smart Visibility Techniques

Virtual windows and transparency can help to direct the user's focus and reduce the density of information displayed. Such visualisation techniques, which provide a clear view of regions of interest, highlight relevant information, and direct the

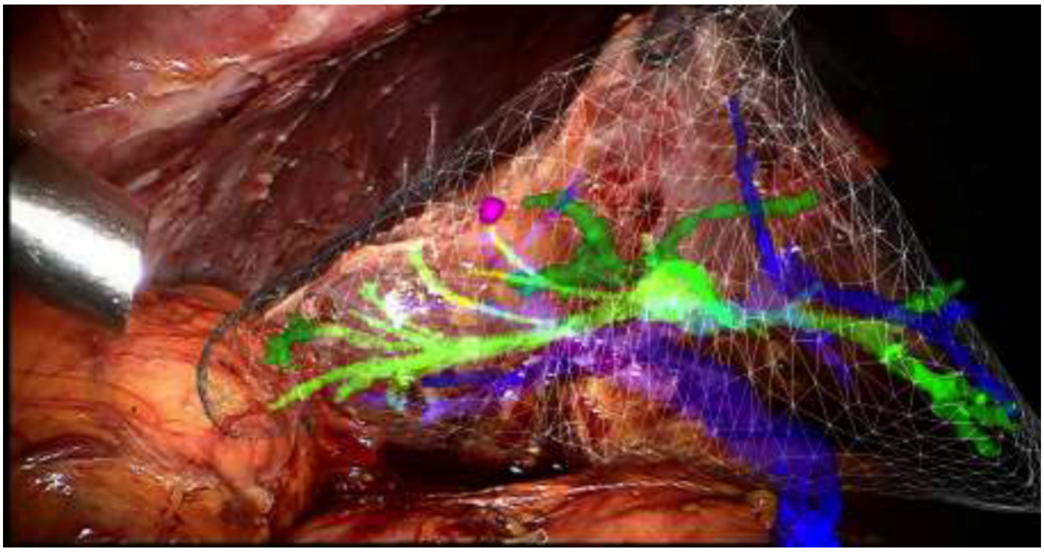


Figure 3.17: Combination of a wireframe and opaque blood vessels on a laparoscopic video feed. The wireframe adds context to the other virtual objects and reduces the floating effect, thus improving depth perception [128]. ©2013 IEEE

Figure 3.18: Tumour and blood vessels during a laparoscopic liver resection. Left: virtual elements seem to be in front of the liver; right: semi-transparency by alpha blending improves depth perception [129]. Figure 1c) and 1d) from [129] excluded due to missing copyrights.

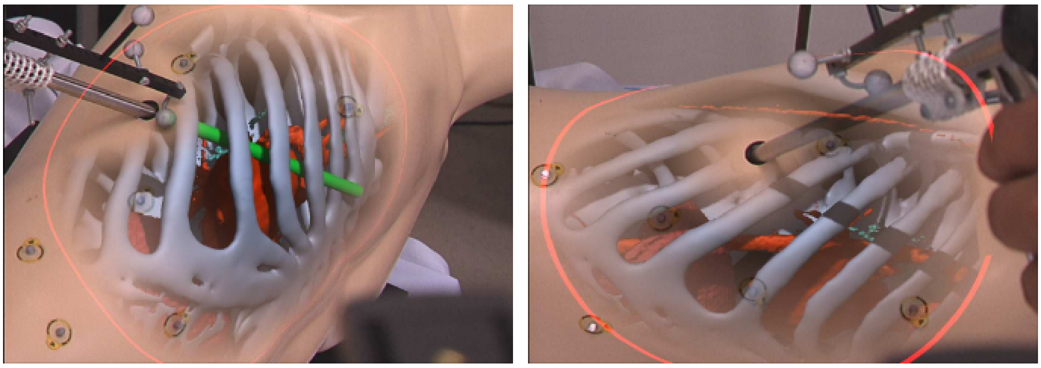


Figure 3.19: The focus and context visualisation forms a virtual window. When moving, the perspective changes and allows for better spatial perception through motion parallax [124]. ©2007 IEEE

user's attention, are called smart visibility techniques [113]. In principle, there are three ways to achieve smart visibility. On the one hand, the visual attributes of the objects can be changed. Hiding elements can be drawn transparently, while signal colours clarify the position of the region of interest. In addition, the geometric objects themselves may be changed. By removing or bending obscuring objects, the view onto the regions of interest can be uncovered. Obscuring elements can also be spatially shifted so that they can still be viewed and the objects of interest are no longer obscured [113].

Approaches to the implementation of smart visibility can be found in technical and medical illustrations. Cutaways remove occluding components of the scene and deform them so that they are perceived as an incision. Ghosted views show coverings transparent in some places and exploded views show the individual components of objects in a form moved apart. These approaches can be applied directly to the representation of virtual data sets [133].

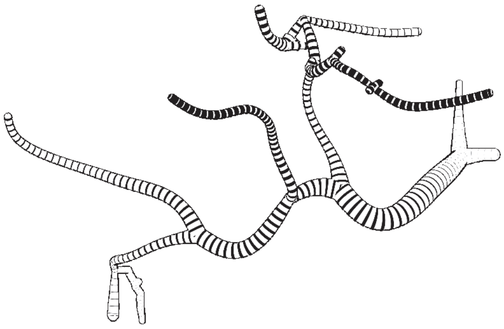
Another method to achieve smart visibility is focus and context visualisation. The objects of interest are called focus while the embedding environment is the context of the focus. This helps setting the objects of interest into relation with the environment. In [134], the context is rendered transparently in relation to the distance to the focus – the further away, the more opaque is the environment. Bichlmeier et al. combined this approach with a virtual window (see Figure 3.19) [124]. The context layer of the virtual content is determined by the outer surface of the objects being viewed; the focus layer consists of elements that are located inside the objects. Areas of the context close to the focus are displayed transparently and thus form a virtual window. The transparency value of the context plane is calculated for each point depending on the respective surface curvature, the angle between the viewing point and surface normal and the distance to the focus plane from the viewer's point of view. By changing the viewing direction and position, other context areas are shown opaquely or transparently, respectively. When fixing the virtual window kinetic depth indications are given by moving the viewing position.

Illustrative Augmented Reality Visualisation

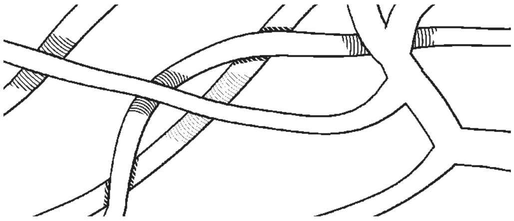
The visualisation approaches presented previously attempt to obtain a spatial impression of virtual elements through realistic lighting models as well as volume and surface rendering. The objects are mostly opaque and thus occlude components of reality, which is often undesired during medical interventions, because the users always need a good view of the patient. Therefore, visualisation techniques with thinner drawings, which conceal little real content, can be useful [113]. Wireframes, as proposed by Sielhorst et al. [132], may be perceived as too confusing due to the high number of elements.

Illustrative visualisations can be a viable alternative to wireframes by providing clearer structures. For example, silhouettes can be used to clarify shapes as well as illustrative texturing, such as stippling or hatching, to encode surface information [113]. A scenario with a high density of information is the visualisation of blood vessels. Here, illustrative renderings can increase the depth perception without cluttering the scene. Ritter et al. [135] proposed a method to represent the shape and course of vessels by hatching lines along the direction of curvature. In order not to use too many lines, the thickness of the lines depends on the angle of the vessels to the viewing position. Furthermore, they recommend to illustrate the overlapping of individual vessels with illustrative shadows. Depending on the proximity of the overlapping vessels to each other, the shadow will be larger or smaller. The distance of individual vessels to the user is encoded by varying the line width of the hatching lines (see Figure 3.20a and 3.20b). Such texturing can already be sufficient to clarify form and topology of vessels. The visualisation is independent of the chosen colour, whereby colour might be used to encode other information [135].

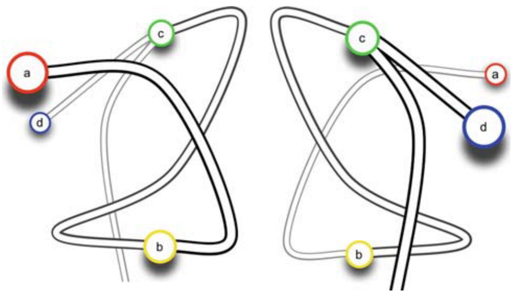
This approach was further investigated by Hansen et al. [136] by encoding depth into the line thickness of silhouettes (see Figure 3.20c). The closer the viewer is to the structure, the thicker the specific part of the silhouette is drawn. Additionally, the hatching technique proposed by Ritter et al. [135] is used to convey distance between vessels and other regions of interest. The combined approach is used to augment the liver in an open liver surgery with anatomical models of tumours and vessels. A comparison between volume-rendering-based and illustrative visualisation is shown in Figure 3.21. Lawonn et al. [137] added to the hatching technique a cylinder and anchors reaching from structures of interest to the cylinder to support the spatial perception. This is illustrated in Figure 3.20d.



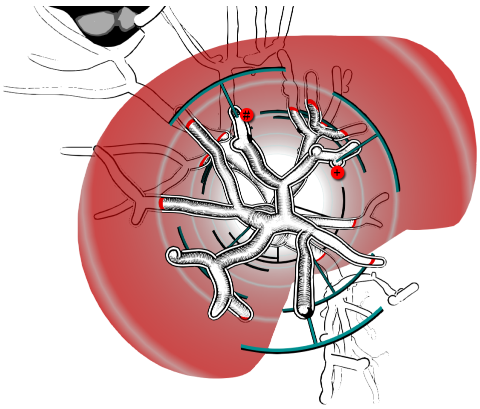
(a) Illustrative depth encoding. Line thickness depends on distance to viewer [135]. ©2006 IEEE



(b) Texturing at intersections indicate distance between the vessels [135]. ©2006 IEEE



(c) Depth encoded silhouettes in two different orientations [136]. ©2006 Springer Nature



(d) Anchors connecting a cutting cylinder and structures of interest [137]. ©2017 Elsevier

Figure 3.20: Illustrative rendering of vessels to improve depth perception.

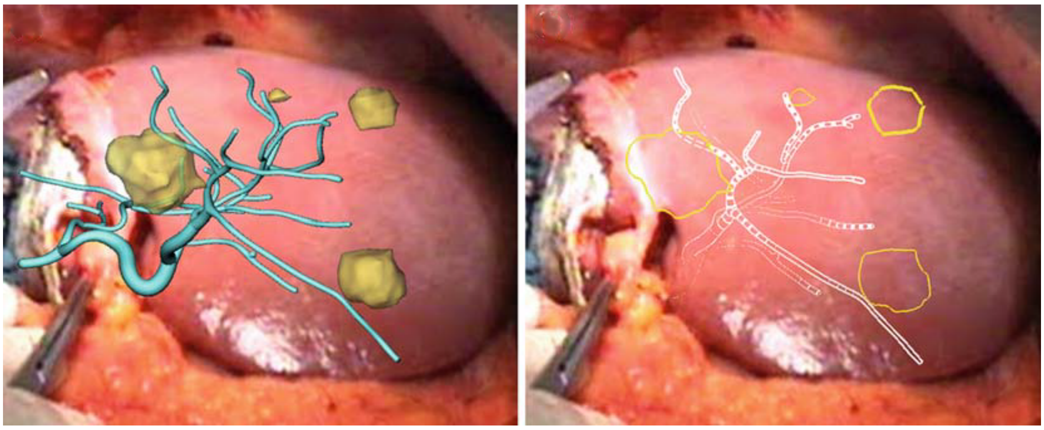


Figure 3.21: Comparison between volume rendering (left) and illustrative visualisation (right) of vessels on the liver in an AR scene [136]. ©2006 Springer Nature

A comparison study showed that the illustrative rendering approach is superior to simple Phong [138] shading and pseudo-chromadepth.

When it comes to projector-based AR, illustrative rendering comes with several advantages over volume and surface rendering. As depicted in Section 3.2.4, projectors are not capable of projecting dark images and visualisation of light content on a light background is difficult. In general, projections are only added to the projection surface which therefore remains always visible. This leads to colours being mixed and the virtual contents appearing transparent. Illustrative rendering provides a very strong contrast and can provide depth cues where the other techniques lack details. This is shown in Figure 3.22.

3.6 Interventional Touchless Gesture Interaction

During medical interventions, especially image-guided interventions, the physician regularly needs to consult the patient's anatomical data. In order to get the correct view of the image data, it must be manipulated by means of zooming, choosing the right image in a set of 2D images, selecting an object, pointing, or rotating a 3D representation of the anatomy.

When it comes to basic 3D interaction, complex interaction often comprises simpler interaction types [139]. According to LaViola Jr et al. [140], these are manipulation, travel, and system control. The most important interaction with medical images is the manipulation of 3D objects and 2D elements. Basic manipulation tasks allow the user to do that which is not possible in the real world.

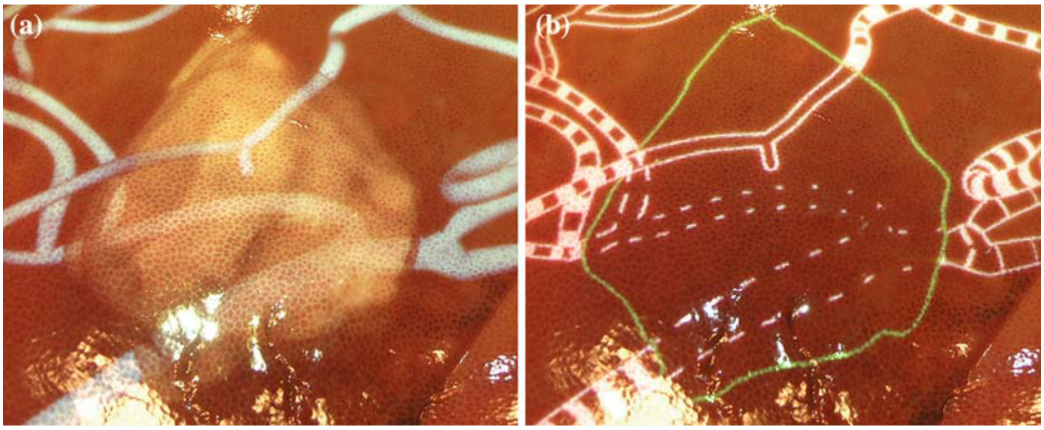


Figure 3.22: Comparison of projections of volume rendered (left) and illustrative visualisation (right) onto a liver. The colours of the volume rendering are mixed with the liver's colour and appear transparent. In the illustrative approach, the details are clearly distinguishable and depth information is preserved [136]. ©2006 Springer Nature

These are selection, rotation, positioning, and scaling [140, p. 258]. Therefore, the interaction with medical images mainly consists of basic interaction tasks.

When performing an intervention, the physician is restricted to sterile interaction in order not to endanger the patient's health with bacterial contamination. However, the interaction is usually carried out with conventional interaction devices, such as joystick, trackball, hardware buttons, touchscreens, mouse, or keyboard. These devices are sometimes not only unsuitable for the type of interaction, they must also be covered with sterile drape to maintain sterility restricting the usability. Touchless 3D UIs can help reducing the infection risk and provide a proper amount of DOF for complex interaction, such as 3D rotation. It should be noted that 3D UIs are not limited to a 3D representation of objects, but rather describe a three-dimensional interaction within the given context [140, p. 7]. The whole topic of 3D UIs is elaborated in great detail in Bowman et al. [141] and LaViola Jr et al. [140]. These books cover information about input and output hardware technologies, human factors, 3D interaction techniques, and general guidelines for designing and developing 3D UIs. Most of these information are relevant for this thesis in various ways, but describing them here exceeds the scope of this dissertation. Therefore, the reader is kindly referred to these standard works for in-depth views on 3D UIs. The same applies to the concept of NUIs, which is explained comprehensively in Wigdor et al. [142] and briefly explained in the following.

3.6.1 Natural User Interfaces

According to design guidelines postulated by Wigdor et al. [142], an NUI is a UI that feels natural to the user, but does not necessarily have to be intrinsically natural. NUIs are easy and quick to learn, memorise, and well-usable with little cognitive workload regarding the interaction. It is therefore important that the user experience feels like an extension of the body for a novice as well as an expert user. The interaction must thus be authentic to the medium and not try to mimic the real world. The NUI considers the context of the application by including the right metaphors, visual indications, feedback, and input/output method for this context. Therefore, it is often not feasible to translate the UI from one genre of computing to another. In order to create a spatial NUI it is crucial to make use of all three axes on input and output [142]. A 3D rotation is naturally performed with a gesture with at least 3 DOF.

The design goal of an NUI is strong immersion of the users such that they no longer compare their actions to a defined pattern. Therefore, feedback is necessary to respond to every input in order to exterminate the feeling of malfunctions. This feedback must be immediate, because a delay would otherwise directly separate the real and the virtual. Major state changes should always be visible, such as the start or end of (possible) input. This is contrary to systems embedding functions in menus. Users should be allowed to manipulate content directly and not through a menu interface, e.g. by using a zoom gesture instead of a button [142].

3.6.2 Gestures

Gestures have been used for long for all kinds of virtual environments and other 3D environments. They may create an illusion of interacting with a virtual environment as if one is not using any input device at all [140]. Hence, gesture interfaces are an integral part of perceptual user interfaces [143] or NUIs [142]. To create a truly well performing and easy-to-learn NUI is challenging. While gesture-based interfaces exist for simple task sets replicating real world actions, such as the Nintendo Wii (Nintendo Co., Ltd. , Kyōto, Japan) or the Microsoft Kinect, more complex gesture interfaces are hard to design [140].

Gestural commands can be classified as postures and gestures. A posture is a static configuration of the hand, whereas gestures are a movement of the hand, probably while being held in a specific posture. The usability of postures and gestures depends on the complexity and number of the commands, because these are proportional to the learning effort of the user [140].

Different types of gestures exist that humans use [140, 144, 145]:

Mimic gestures Gestures that are used to describe a concept, but are not connected to speech.

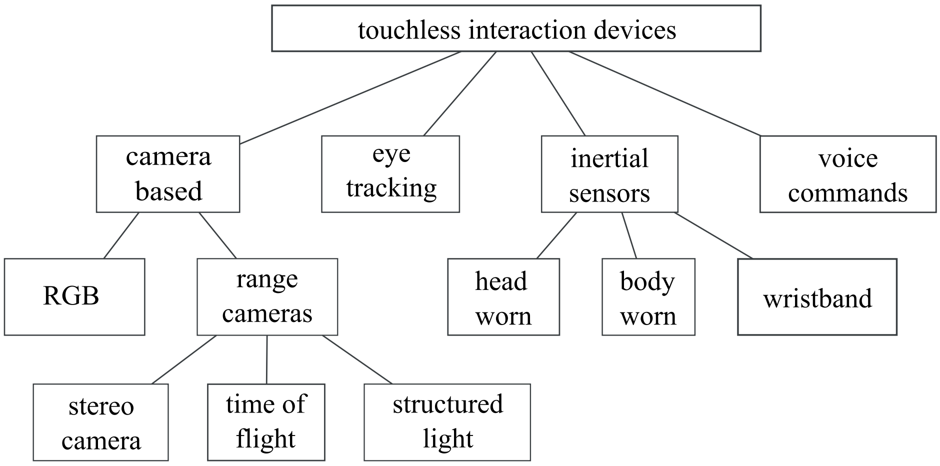


Figure 3.23: Different modalities to recognise touchless input during interventions.

Symbolic gestures Gestures to express things like insults or praise, e.g. "thumbs up".

Sweeping Gestures used for marking-menu techniques. These are menus in pie-like menu structure.

Sign language The use of a set of postures and gestures in communication with hearing-impaired people or finger counting.

Speech-connected hand gestures Spontaneous gestures performed unintentionally during speech or language-like gestures integrated in the speech performance (sometimes referred to as metaphoric). A special type is the deictic (pointing) gesture used to indicate a referent (object or direction) during speech. These have been studied intensely in human-computer interaction (HCI).

Surface-based gestures Gestures made on multi-touch surfaces. These are 2D, but can be used for 3D systems in hybrid interfaces.

Whole-body gestures Instead of only using hand and arm movements, feet or the whole body may be used. They can be mimic or symbolic gestures.

3.6.3 Devices for Touchless Gesture Recognition

To recognise touchless input for interventional use, different types of devices have been used for different purposes [146]. Figure 3.23 gives an overview. Although voice commands and eye tracking tracking were used to log anaesthesia records [147], trigger functions of an integrated operating room (OR) [148], and support telerobotic interventions [149–151], these modalities are not discussed in this thesis due to its focus on finger, hand, and body gestures to control medical image viewers and other interventional software. The aforementioned input modalities on their own are either not precise enough or not suitable for these tasks [146], but might be beneficial in a multimodal approach as proposed by Hatscher et al. [152, 153]. In addition, inertial or myoelectric sensors can be used for gesture recognition.

To recognise hand and body gestures for interventional control of software, mainly camera-based devices are used. These are either single RGB-cameras or range cameras, such as stereo, time-of-flight, or structured-light cameras. With an RGB camera, Wachs et al. [154] detected hand gestures and postures by segmenting the hand in the image based on its colour from the background. The difference between two consecutive frames is computed and serves as a motion cue that can then be interpreted as a gesture.

In research, new interaction concepts are often evaluated with off-the-shelf devices designed for entertainment applications. With the introduction of the first Microsoft Kinect, a structured-light-based range camera to generate in real time a 3D representation of the environment within the FOV and recognise whole body gestures, many research groups explored the device in the medical context [155–167]. The device projects an infrared light pattern that is used to generate a depth map at a rate of 9 to 30 Hz and at a distance of 1.2 to 3 m (see Figure 3.24). In terms of gesture recognition the device is, however, limited to recognising the posture of the arms or the body.

Soutschek et al. [169] used a time-of-flight-based depth sensor as a similar 3D reconstruction technology to recognise arm and hand gestures. By applying a threshold to the depth data a coarse segmentation of the arm is performed. The hand is separated from the background based on the colour. This technology was later commercially distributed with the Microsoft Kinect 2 that enabled the tracking of the position and orientation of 25 body joints including the thumbs. Presumably because the new Kinect did not add any significant new tracking capabilities compared to the first Kinect it was only adopted by few research groups [170–172].

Another off-the-shelf gesture recognition device, the Leap Motion Controller (LMC), has been used in the interventional context [173–181], but in contrast to the Kinect it enables the tracking of the hands and fingers. The tracking frequency is approximately 100 Hz at sub-millimetre precision. Technically the LMC is a stereo camera, i.e. two cameras calibrated in such a way that it is possible to generate a disparity map and derive a distance for each pixel in the image. Gestures can be performed within a volume of approximately $0.5 \times 0.5 \times 0.5\ m$. The same principle has been explored before by Kipshagen et al. [182], who used a stereo camera to recognise hand postures and gestures to control a medical image viewer (see Figure 3.25).

Figure 3.24: A physician performs arm gestures in front of a screen and the Microsoft Kinect [168]. ©2014 Kenton O'Hara, Microsoft Research

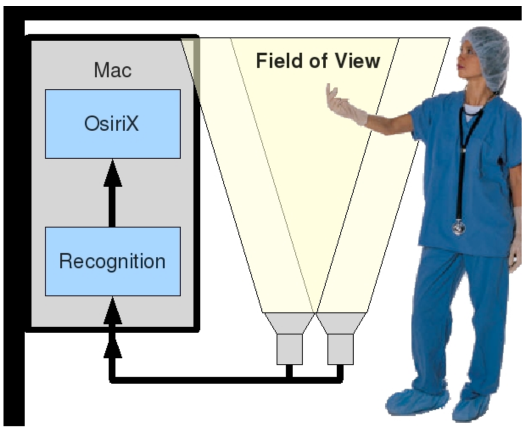


Figure 3.25: Schematic drawing of a physician performing hand gestures above a stereo camera [182]. ©2009 Springer Nature

Figure 3.26: A physician interacting with the interventional software via the Myo armband [171].

To track arm gestures, instead of a camera also inertial sensors may be used that are distributed on the body [183, 184] or integrated in a wristband [185]. These inertial sensors consist of an accelerometer, magnetometer, and gyroscope to track the user's relative movements. In order to distinguish the gestures from other movements, a lock/unlock trigger is often necessary. Because of the drift that adds to the relative movement due to inaccuracies when calculating the position from measured accelerations, inertial sensors should be used in combination with other tracking modalities by means of sensor fusion.

Instead of inertial sensors, such wrist- or armbands can also be equipped with myoelectric sensors. These gather information on finger and hand movement by measuring the electrical signals of contracting muscles. A commercially available version of such a myoelectric armband was the Myo armband (Thalmic Labs, Inc, Kitchener, Canada). It provided an SDK with pre-defined gestures to be mapped onto software functions and could be worn under the clothes, which was especially interesting for sterile interventional scenarios. However, the device had a high false positive recognition rate making special lock triggers necessary [171, 186].

3.6.4 Challenges in Gesture Design

The design of a gesture set depends heavily on the capabilities of the input device being used. Not only must the hands, fingers, or other body parts be tracked at a low level. These tracking data must also be translated into motions that are recognised as gestures at a high level. This is a complex process of using machine learning or heuristics that is still not completely reliable. Thus, gesture input devices often generate noisy tracking data leading to inaccurate input and requiring smoothing or calibration. Such a calibration is, however, inadequate in some areas, such as in the public with changing users [140, p. 401]. When accessing menus, the jittering input data may also demand for larger menu items.

Because gestures are not only used to generate input into software but also in everyday life, they need to be clearly distinguished from unintended action [187]. Therefore, they need a clear delimiter indicating the initialisation and termination of a gesture. This issue is called the gesture segmentation problem [140, p. 401]: a gesture must be separated from a constant stream of input data. A possible solution is the push-to-gesture technique that implicitly or explicitly makes use of a mechanism to ensure only intended gestures are captured to be recognised similar to "push-to-talk" with speech interfaces. This can be achieved by enabling gesture input only in a dedicated area or by performing an unlock command. Bigdelou et al. [184] compared the use of a voice command to activate the arm gesture recognition with a handheld switch as unlock device. A study revealed that the users preferred the handheld switch over the voice command, probably because of a faster response time and direct haptic feedback. Jacob et al. [164] used the body orientation relative to the screen the user interacts with as an indicator for the users intent. Only when the user's body faces the screen the gesture recognition is active. Strickland et al. [159] defined a 3D workspace volume in which the arm gestures are to be performed ignoring unintended gestures outside this volume. A similar approach was presented, which defines an area in the camera image in order to trigger the start of the swipe gesture recognition, an area for the swipe gesture to be performed, and an area to deactivate the gesture recognition (see Figure 3.27) [188].

Multimodal approaches should also be considered, as they increase the usability if applied appropriately. Mentis et al. [189], for example, used voice commands as a mode switcher and function trigger in their medical image viewer, but hand and arm gestures for continuous manipulation actions.

When creating a gesture set for an existing application, the gestures are often mapped onto mouse events or keyboard shortcuts, disregarding the benefits of an NUI (see Section 3.6.1) and not taking advantage of the 3D input capabilities of the users' hands. For example, Mauser et al. [175] used simple hand positioning as gestures and mapped them directly onto mouse events to allow for basic interaction

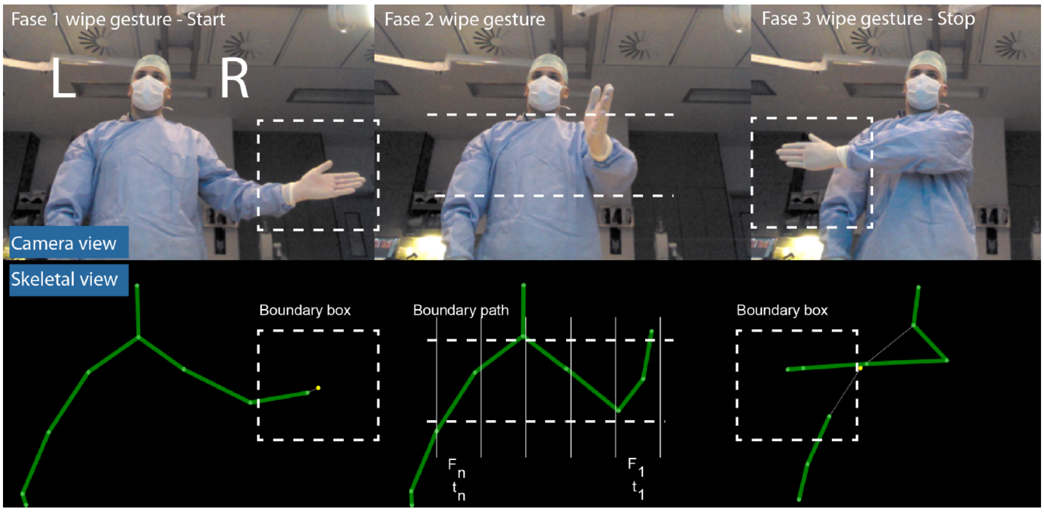


Figure 3.27: Three zones in the image to activate gesture recognition, perform the gesture, and deactivate again [188]. ©2013 IEEE

tasks with medical images. Thus, only the two DOF of the mouse were used, which is unsuitable for 3D rotation.

The gestures available in a UI are typically invisible to the user, so that the gesture set needs to be discovered. Hence, the gestures should be easy to learn. This is especially important for professional domains, such as medical interventions. According to a study on the usefulness of hand-gesture interaction performed by Stevenson et al. [190] simple natural gestures are preferred. The authors performed interviews with surgeons and observed their interventions. The results show that the willingness of the surgeons to invest time and effort into complex interaction varied. Hence, a simpler gesture set is more likely to be accepted by the users, because it is easier to learn.

Depending on the user's level of experience the total number of gestures may vary between novices and expert users. In any case, the cognitive load must be kept at a reasonable level in order not to disturb the user's primary task. To prevent confusion, feedback must be given to the user on the recognised gestures [140, p. 402].

4 Projector-Based Augmented Reality to Support MRI-Guided Interventions

In the context of needle-based percutaneous ablations the radiologist's essential task is to reach the tumour target with the ablation needle precisely on a safe path (see Section 2.2.1). This path is planned in advance based on a planning dataset and combines the optimal accessibility of the tumour with the maximum security: First, the target position is determined. Then an entry point is chosen that ensures sufficient distance to risk structures along the path, such as blood vessels, nerves, or the lung. The radiologist relies on this planning data during the intervention and inserts the ablation needle with as little deviation from the planned path as possible. During MRI-guided interventions, the entry point is often found with the finger tipping method or by using a tape measure (see Section 2.2.2). Then the needle is guided through the skin to the tumour while constantly acquiring images along this path [13, 19] and estimating the correct movements accordingly.

Navigation systems support this process by providing at least the following information:

- the entry point on the patient's skin through which the needle is inserted,
- the target point, i.e. the lesion,
- an orientation aid to help align the needle with the planned path,
- depth cues to indicate the depth of the needle and the distance to the target, respectively.

Often, there is also information on risk structures and their relation to the applicator.

In this chapter, the first applicator guidance system for the use inside the MRI bore is proposed which provides essential applicator navigation information directly on the operating field by means of projector-based AR (see Section 3.2.4). The system is designed for percutaneous liver tumour ablations with the radiologist performing the intervention from the feet side of the MRI in a static scenario without organ or body surface deformation, but may also be used for other needle-based MRI-guided applications. The goal is to decrease correction movements with the applicator that are sometimes necessary if it is not on the planned path anymore by providing navigation cues that are accurate and easy-to-use. To this extent, two navigation visualisation concepts are presented that aim to assist the radiologist while targeting the lesion.

First, a short overview on already existing applicator navigation support approaches in the MRI and on projector-based AR intervention support is given. Second, the hardware setup and its calibration are explained as well as the navigation concepts. The latter include information about the entry point and target position as well as needle alignment support and depth cues. Finally, the results of the user study assessing accuracy, usability, and subjective mental demand of the navigation system are presented and discussed afterwards.

4.1 Related Work

In order to safely and effectively guide instruments to a target lesion, appropriate instrument guidance is essential to simplify and shorten the intervention. The support of needle guidance in image-guided procedures is an active area of research.

Many of the proposed systems divide the needle puncture into the three subtasks tip positioning at the entry point, needle alignment to the planned path, and needle insertion to the target, as proposed by Seitel et al. [191]. Information about these subtasks is presented to the radiologist during the intervention. The kind of guidance depends on the current subtask. For example, a crosshair visualisation for needle positioning and alignment was introduced, supported by a progress bar indicating the needle depth [191, 192]. A video stream is superimposed with the preview visualisation of the needle in the planned pose to which the user aligns the applicator in order to reach the target [193].

When using tracked instruments that are calibrated with the MRI, the real-time imaging can be aligned along the needle to always get the view about the needle's surroundings, as shown by Kägebein et al. [71, 194]. Basic navigation cues in the form of the planned needle path are also provided. This approach is promising, because it respects the requirements of the interventional MRI and is up to date with deformed inner structures without the need for registration. However, this system lacks a way to quickly find the needle entry point and the needle tracking volume is small.

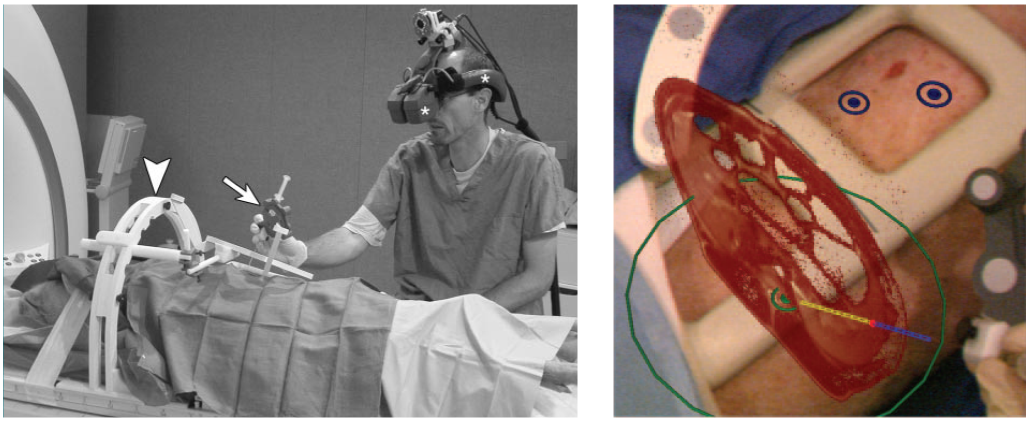


Figure 4.1: A video see-through AR HMD is used to overlay the patient with 2D images used by Wacker et al. [203]. Left: A radiologist wearing the HMD with a stereo camera and an infrared camera attached. Right: Visualisation of a 2D images and a virtual needle. ©2006 RSNA

Navigation cues to support instrument guidance are often displayed on a screen [193] that separates the virtual information from the patient and increases mental stress [26]. To solve this problem, AR can be used to merge virtual guidance information with the real scene the radiologist looks at. AR navigation systems were already successfully set up using a microscope for spine surgery [195], in laparoscopic surgery [196, 197], plastic surgery [198], or neurosurgery [199, 200] among many other scenarios [201, 202].

The first AR system to support MRI-guided interventions was presented by Wacker et al. [203]. They augmented the patient with two-dimensional anatomical images and a virtual representation of a tracked needle in a video see-through HMD (see Section 3.2.3) to support biopsies (see Figure 4.1). The HMD itself has been previously introduced by Sauer et al. [204] and Das et al. [88] for use during CT interventions. A stereo camera attached to the HMD enables stereo vision for the radiologist. Markers are placed on the biopsy needle and a stereotactic frame which is relative to the patient. Those markers are reported to be accurately tracked with an infrared camera also placed on the HMD [88, 203]. The two-dimensional image is fixed to the transversal plane. However, due to the strong magnetic field of the MRI scanner, the HMD cannot be used close to the MRI. Therefore, the patient must be translated out of the bore after image acquisition which does not allow for real-time imaging during the puncture. Hence, the radiologist must rely on the planning data set which is out of date as soon as the patient breathes or moves. Additionally, the fixed orientation of the image slices limits the needle path to be placed in the transversal plane, which is often not feasible depending on the location of the lesion. Besides that, the radiologist is not able to see the real scene

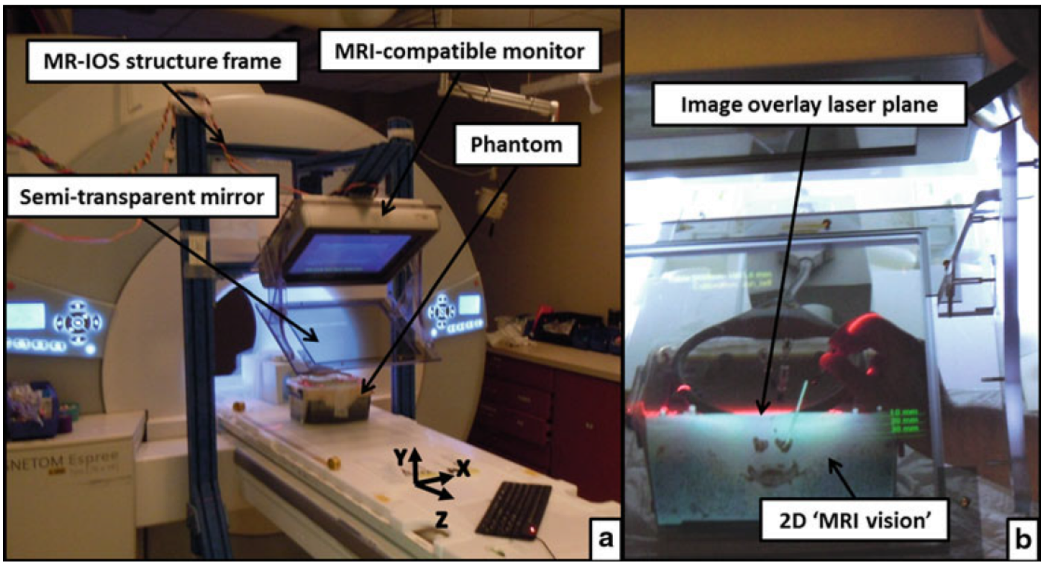


Figure 4.2: An optical see-through mirror is placed by [207] in front of the MRI bore to present 2D MR images aligned with a phantom [207]. ©2012 Springer Nature

directly, but only on the display. This limits the FOV and resolution of the view, which is not desired in medical settings[1].

Another approach on AR to support MRI-guided interventions was followed by Weiss et al. [205] who adapted the optical see-through system (see Section 3.2.1) first presented for CT interventions [206] to be used at the MRI (see Figure [207]). As it can be seen in Figure 4.2, the system is placed above the MRI's patient table directly in front of the bore. It is designed to be used with the in & out technique, i.e. the patient is moved out of the bore to perform the intervention and into the bore to acquire image data. The optical see-through system and the MRI are registered such that a single transversal plane is overlaid on the operating field marked with a line laser together with a virtual needle illustrating the current needle position. This setup has several drawbacks. The size of the construction restricts the already limited patient access even more. Further, the possibility of showing only transversally oriented planes strongly limits the planning of the needle path to this orientation, which does not comply with the workflow of MRI-guided interventions. The most important drawback is the need to translate the patient out of the bore for the system to work, neglecting most of the MRI's advantages, such as real-time imaging and thermometric therapy control.

[1]Exceptions are laparoscopic or telerobotic interventions, where the advantages of the type of intervention outweigh the limitations a video stream of the operating field comes with.

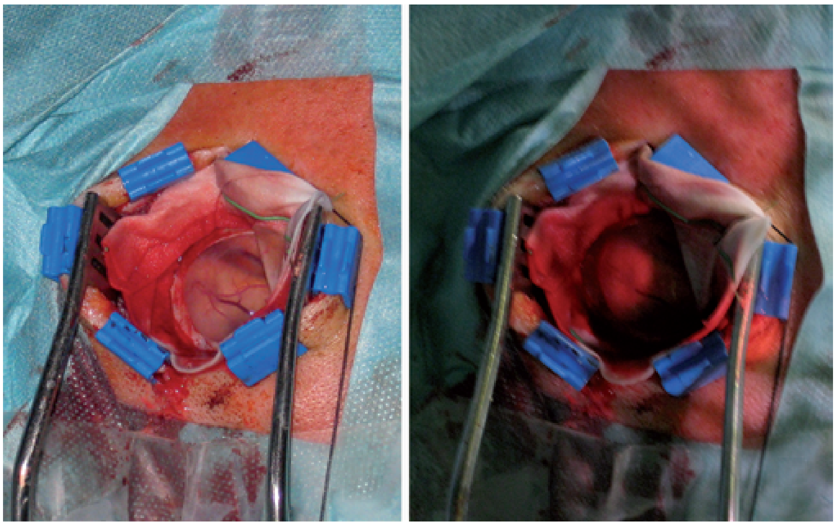


Figure 4.3: Projection of a tumour model onto the brain. Left: the tumour cannot be distinguished from the brain; right: the projection helps to find the tumour location [199].

As the workspace is limited and the use of materials is restricted due to the strong magnetic field, large and/or MRI-unsafe devices such as HMDs, monitors, or mirrors in front of the physician cannot be used in-bore. However, projector-based AR suits these requirements. There have already been several approaches to support minimally-invasive interventions with projected AR. Sugimoto et al. [208] overlaid the patient with anatomical images during gastrointestinal, hepatobiliary, and pancreatic surgery. Nevertheless, this was rather a proof of concept of markerless surface registration than a support of the actual intervention, because no navigation cues were given and the instruments were not included in the overlay. Besharati Tabrizi et al. [199] augmented the visible portion of the brain with a model of a tumour that is otherwise only barely distinguishable during neurosurgery to facilitate its location (see Figure 4.3). A projector-based approach to guide a needle to a tumour target in the liver was taken by Gavaghan et al. [209], who integrated a tracked handheld projector into a navigation system used in liver surgery (see Figure 4.4). The projector enables the physician to flexibly augment the operating field with anatomical information or navigation cues [79, 210]. The projected navigation hints include the aforementioned crosshair visualisation and a needle path preview (see Figure 4.5). The advantages of this system setup are its size and flexibility; the projector can always be placed such that it illuminates the operating field without a line of sight problem, which is valuable in environments with poor patient access. However, the projector needs to be held in the hand – probably by an assistant – and cannot easily be made MRI-compatible. In alternative to such a

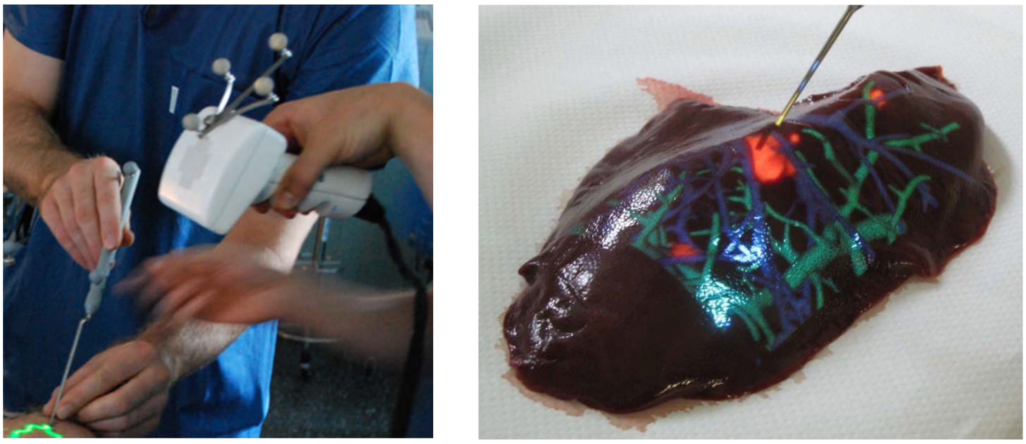


Figure 4.4: The handheld projector AR navigation proposed by Gavaghan et al. [209]. Left: the projector is pointed towards the operating field; right: the projection of anatomical objects and the virtual extension of a needle on a liver. ©2012 Springer Nature

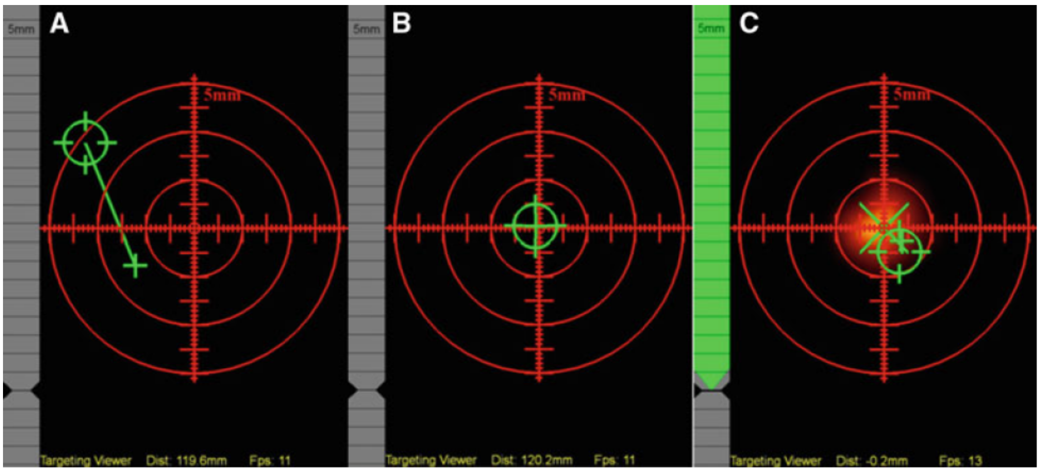


Figure 4.5: Crosshair navigation aid with depth-indicating progress bar and virtual needle to support needle punctures [210]. ©2012 Springer Nature

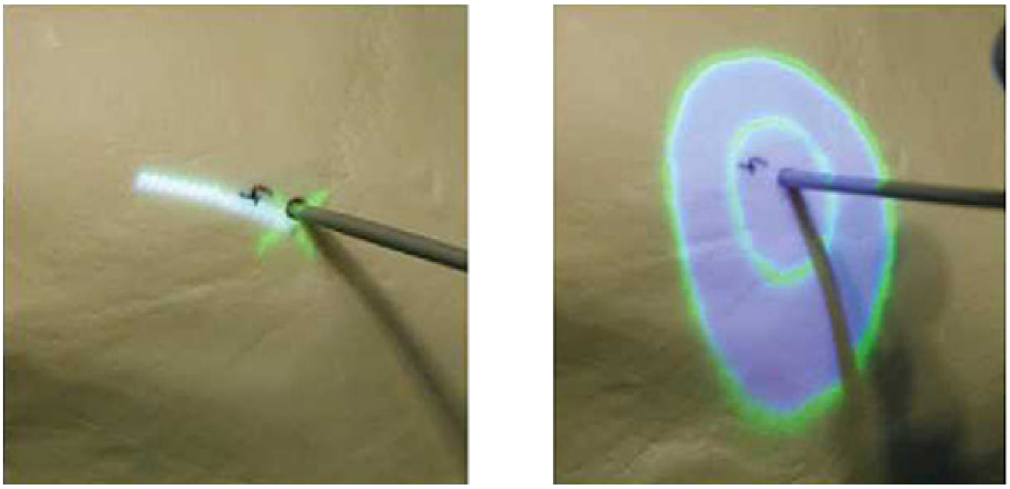


Figure 4.6: Projected explicit navigation cues proposed by Krempien et al. [211]. Left: the entry point is marked with a cross, an arrow directs the needle to the correct orientation; right: the inner circle expands towards the outer circle when approaching the target. ©2008 Elsevier

crosshair visualisation, depth cues can also be given as concentric circles, with the inner circle expanding as depth increases, as described by Krempien et al. [211]. Here, the needle orientation is guided by an arrow (see Figure 4.6).

In contrast, AR in the form of auditory feedback was proposed for guiding needles [212]. A combination of visual and auditory augmentation was proposed by De Paolis et al. [213] and Bork et al. [214]. Bork et al. combined both acoustic and visual feedback to encode distance information between a surgical instrument and areas of interest. Their method includes a spreading shape around the instrument tip that increases over time. Acoustic signals underline the growth process and speed of this shape. In an MRI-guided intervention, however, acoustic feedback is only of limited use due to the noise of the scanner.

As it becomes clear, none of the existing approaches except the automatic needle-pose-dependent image alignment approach by Kägebein et al. [71] is able to meet the strict environmental requirements of the MRI and integrate well into the workflow of MRI-guided interventions. Therefore, in this chapter, a new approach of projected AR for in-bore usage to support MRI-guided interventions is presented that addresses these issues.

4.2 Materials and Methods

To realise a spatial AR environment with a projection inside an MRI scanner, align image data to the patient or phantom, and visualise it in the operating area, the following steps are necessary:

- Hardware Setup
  - Placing the projector and adjusting zoom and focus.
  - Positioning a camera to view the whole projection area.
- Calibration Process
  - Calibrating the projector with the camera.
  - Registering the MRI coordinate system with the projector-camera system.
- Surface Reconstruction
  - Generating a virtual point cloud representation of the projection surface with a structured light approach.
  - Generating an anatomical patient dataset with the MRI.
  - Segmenting and meshing the structures of interest.
- Visualisation
  - Determining correspondence between 3D points of the surface point cloud and the world coordinates of the projector's pixel positions.
  - Projecting 3D patient data and navigation clues, physically aligned and perspective-correct.

These steps are described in detail in the following sections.

4.2.1 Hardware Setup

A Siemens MAGNETOM Skyra 3T MRI (Siemens Healthcare AG, Munich, Germany) with a 70 cm bore was used. The strong magnetic field must be considered when designing a system of hardware components to work inside the bore. For prototyping, all hardware used inside the 5 Gauss line[2] of the MRI must at least meet the requirements that the noise introduced by new hardware must not

[2]The 5 Gauss line marks the distance from the MRI scanner within which the static magnetic field is stronger than 5 Gauss (0.5 mT). This is the highest allowed field for implanted cardiac pacemakers.

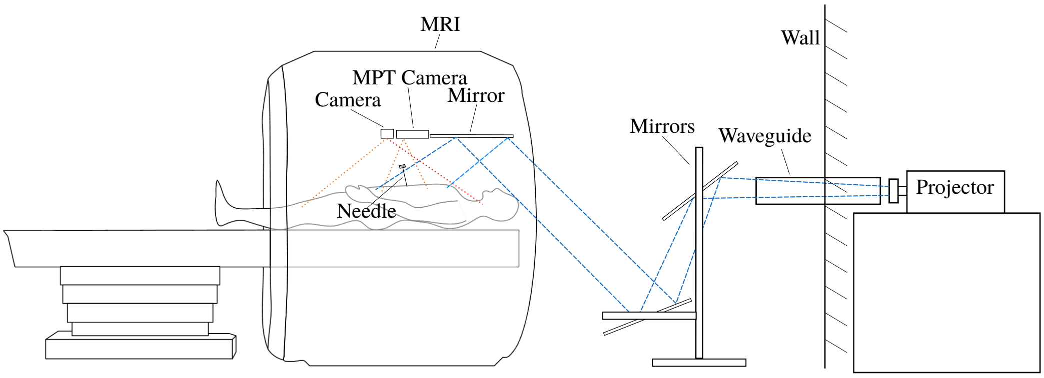


Figure 4.7: Schematic drawing of the complete hardware setup. ©2018 John Wiley and Sons

significantly affect the imaging, and, more importantly, that it must not endanger people by forces caused by the static magnetic field. This is best achieved by placing as few components as possible inside the 5 Gauss line and shimming the remaining parts. For medical products, this is not sufficient as they have to comply with MRI safety standards [215].

Referring to Sections 3.2.4 and 3.3.2, the minimum hardware needed to create a projector-based AR is a projector and a camera to calibrate it. Powerful projectors with high contrast are often bulky and include a power supply unit that is magnetic and cannot easily be replaced by a battery. Therefore, the projector should be placed at a remote location. To be able to project over the long distance, decent optics must be included.

Projector

An NEC PX700W Digital Light Processing ultra-long-throw projector (NEC Display Solutions Europe GmbH, Munich, Germany) was placed outside the MRI room on the head side (see Figure 4.8a). Because in this setup the 5 Gauss line coincided with the scanner room's walls, the projector was located behind the wall. Its light is guided through a waveguide and three mirrors (see Figure 4.8b) onto the operating field (see Figure 4.7). The in-bore mirror was mounted horizontally at the ceiling to use as little space as possible (see Figures 4.8c and 4.8d). This leads to a small incidence angle such that the image is stretched in the height dimension. The projector's resolution was decreased to 1024×768 px to reduce the image dimensions so that it fits the mirrors' sizes and all projector pixels can be observed by the camera during the calibration process. This results in a projection distance of 3.5 m and a real image size of 35×47 cm (width by height). The built-in keystone correction is not applied, because the calibration algorithm calculates coefficients for all occurring distortions. Otherwise, the stereo calibration would be corrupted due to false projector intrinsics.

Camera

In order to calibrate the projector a camera is needed (see Section 3.3.2). For the best possible calibration result, the camera needs to see all projected pixels. Hence, a Qumox SJ4000 wide-angle camera (Qumox, Kowloon Bay, Hong Kong) with a 170° opening angle is placed on the bore ceiling right above the projection inside the bore next to the isocentre, where the magnetic field is homogeneous. This way, there is no momentum applied to the camera and it remains in place. To further reduce the magnetic effect on the camera, the USB port powering the device was removed and the USB line was soldered directly onto the circuit plate. The camera works at a resolution of 1280 ×720 px with 30 FPS from a distance of 20 to 47 cm to the projection area.

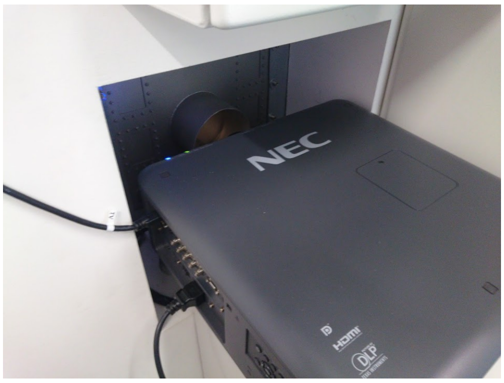
In preliminary tests the positioning and correct functioning of the camera was evaluated. An endurance test revealed that the camera was held in place by Velcro tape for the whole duration of two days. While scanning with different T1- and T2-weighted sequences the camera remained functional. Image acquisition with the MRI was only little affected by noise from the camera because there were only very few metallic parts left. However, when observing the projected image, a strong rainbow effect caused by the projector's rotating colour wheel became apparent. This was solved by increasing the camera's exposure value to EV +1 and the film speed to ISO 100.

Needle Tracking

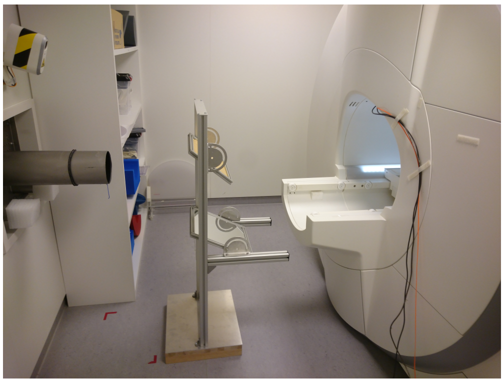
For the setup of a needle navigation test scenario, an MRI-conditional MPT camera (Metria Innovation, Inc., Wauwatosa, WI, USA) was mounted inside the bore, next to the mirror and wide-angle camera, to track MP markers (see Section 3.4.1) attached to the needle instrument. The pose of the markers are obtained with a position error of less than 1 mm at a distance of 2.5 m and an orientation error of 0.05°. The camera's tracking rate is between 1 and 15 FPS, depending on the distance: the nearer the tracked markers, the slower the marker recognition due to the fixed focus plane lying behind the marker. The FOV of the camera is rather small ($10 \times 15\,cm$).

4.2.2 Calibration Process

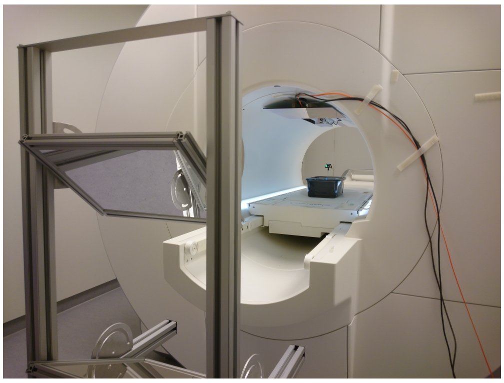
The projector was calibrated with the structured-light approach proposed by Moreno et al. [104], which is explained in Section 3.3.2. A checkerboard of known size



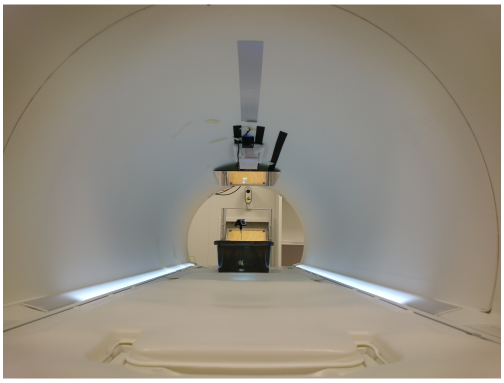
(a) Projector outside the scanner room in front of the waveguide.



(b) The mirror stand and waveguide on the MRI's head side.



(c) View into the bore from head side.



(d) View from feet side showing phantom, needle, MPT, and wide angle camera.

Figure 4.8: Hardware setup in the MRI scanner room.

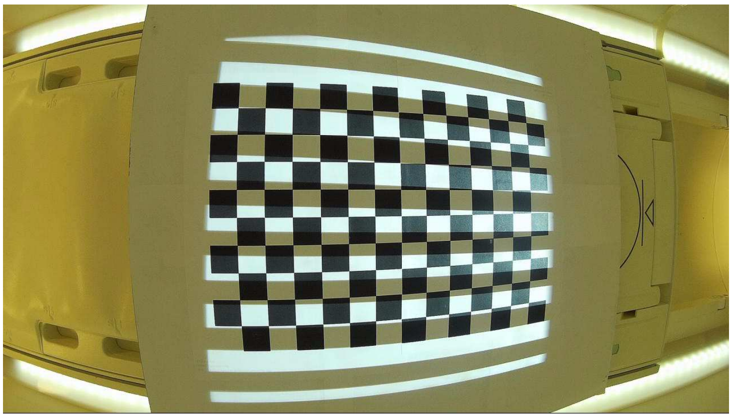


Figure 4.9: One of the projected grey code patterns on the chessboard during projector calibration from the view of the ceiling-mounted wide-angle camera. ©2018 John Wiley and Sons

was placed inside the bore to be covered completely by the projected image. 22 grey code patterns were projected onto a chessboard (see Figure 4.9) and the local homographies, i.e. transforms between the calibration points projected onto the chessboard and those observed by the camera, are calculated. In this manner, correspondence between pixels and world coordinates is determined. The projector and the camera were calibrated as two separate cameras with the help of the OpenCV[3] library. To compute the intrinsics of the wide angle camera the calibration approach of [104] was adapted to use the intrinsics from a separate fisheye camera calibration that uses a ChArUco board [114, 116] (see Section 3.4.1) instead of an ordinary chessboard. This board has more distinguishable features that are used to interpolate between corner points and compensate for errors, which results in a more accurate calibration. The stereo calibration yields the world coordinate transform between the camera and the projector. Six sets of grey code patterns on the chessboard – each in a different orientation – were used for the projector calibration. This calibration step has only to be performed initially or when the projector, camera, or mirrors have been moved.

In order to be able to project the anatomical patient data onto the operating field, the projector-camera system has to be registered with the MRI. Accordingly, the transform $T_{pat,cam}$ between the patient coordinate system in the image data set and the wide-angle camera coordinate system – which serves as the world coordinate system – needs to be found. A calibration body was 3D-printed and filled with transparent liquid candle wax consisting of paraffin, Vaseline, and white oils (see Figure 4.11). It provides a clear contrast in images from a T1 sequence and causes

[3] http://opencv.org/

Figure 4.10: Estimation of the fixed ChArUco marker pose for registration of the MRI coordinate system with the camera coordinate system. ©2018 John Wiley and Sons

little noise. The parallelism of the body's edges in the corresponding MRI dataset is best ensured by the scanner's built-in 3D distortion correction. A ChArUco board is attached on top of the plane surface of the calibration phantom so that it is aligned in parallel with the candle wax. The pose $T_{cam,cb}$ of the board in world coordinates is estimated with the OpenCV function cv::aruco::estimatePoseCharucoBoard() [116].

The corner points of the calibration body in camera coordinates are calculated from the known dimensions of the candle wax body (without the 3D-printed case, i.e. 115 ×115 ×117.5 mm) and the axis-parallel ChArUco marker pose $T_{cam,cb}$. The thickness of the outer walls of the calibration body is taken into account by the transform $T_{cb,p}$.

$$T_{cam,p} = T_{cam,cb} \cdot T_{cb,p} \tag{4.1}$$

The corresponding corner points in patient coordinates are then measured in the DICOM dataset of the calibration body. The transform $T_{p,pat}$ between the corner points in the camera coordinate system and the patient coordinate system was estimated by the RANSAC-based cv::estimateAffine3D() from the OpenCV library with a confidence of 0.99.

$$T_{cam,pat} = T_{cam,p} \cdot T_{p,pat}. \tag{4.2}$$

To save the transforms from this registration, a ChArUco marker is fixed to the bottom of the MRI bore (see Figure 4.10), for which the transform $T_{fix,pat}$ to the patient coordinate system is calculated with

$$T_{fix,pat} = T^{-1}_{cam,fix} \cdot T_{cam,pat} \tag{4.3}$$

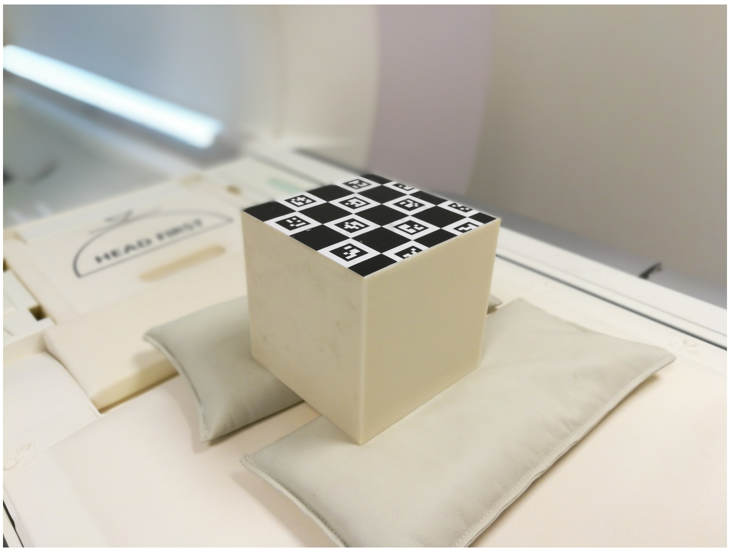


Figure 4.11: Calibration body filled with candle wax for MRI imaging and with a ChArUco board attached to estimate the phantoms pose in camera coordinates. ©2018 John Wiley and Sons

and saved to a calibration file. This way, only the projector calibration needs to be repeated when the camera, projector or mirrors have been moved.

Finally, the MPT system, the puncture needle, and the wide-angle camera need to be calibrated. Four MP markers were attached to the needle and calibrated with a tracked calibration body which led to the transforms $T_{mar,n}$ for every marker. The pre-existing cross-calibration (registration) between the MPT system and the MRI [71] $T_{mpt,mri}$ is read from a calibration file. The MPT system and the wide-angle camera share the MRI as a common coordinate system, which is used for determining the transform. Altogether the transform

$$T_{cam,n} = T_{cam,pat} \cdot T_{mri,pat}^{-1} \cdot T_{mpt,mri}^{-1} \cdot T_{mpt,mar} \cdot T_{mar,n} \tag{4.4}$$

between the wide-angle camera and the needle tip is calculated to be able to visualise the needle in the correct position and orientation later. $T_{mri,pat}$ is the transform between the MRI device coordinates and the patient coordinates depending on how the patient is positioned. This one-time calibration step takes about 30 min, provided that the hardware is already installed.

4.2.3 Surface Reconstruction and Visualisation

Before anatomical data and navigation cues can be projected onto the operating field, the projection surface needs to be transferred into a dense point cloud via a structured light method.

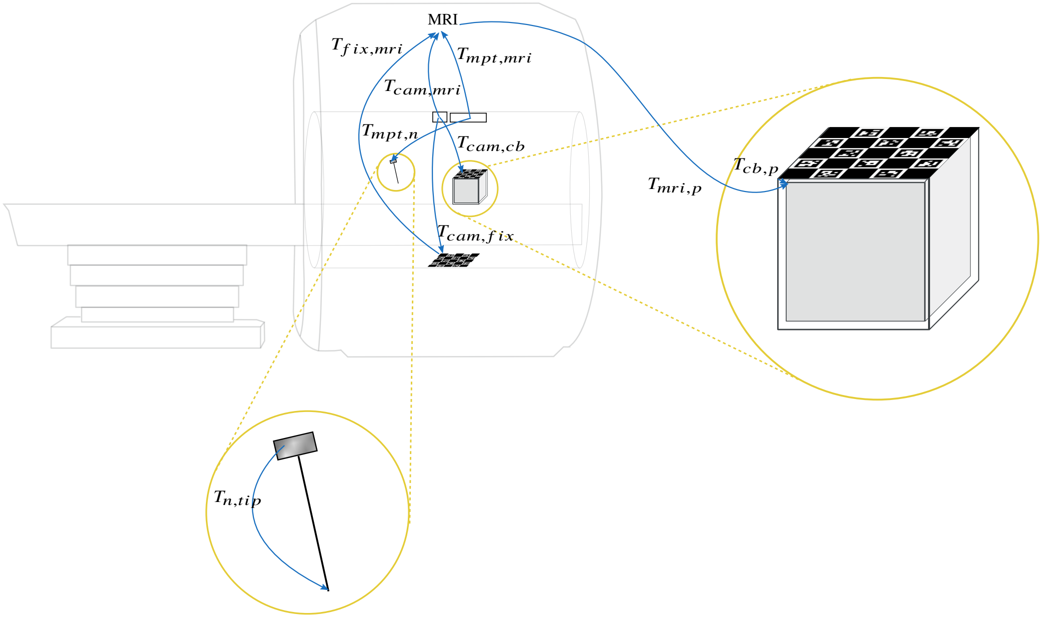


Figure 4.12: The different transforms involved in the registration process.

When all correspondences between a camera and projector are known (see Section 3.3.2), the projector-camera system can be used to 3D scan the projection surface. The result of the 3D scan is a point cloud, i.e. a set of voxels in world coordinates representing the surface. The single points are obtained via triangulation: First, a set of grey code pattern is projected similarly to the calibration step. The projected pattern gets distorted by the uneven surface, so that the camera pixels are now matching different projector pixels during the decoding step compared to the calibration process. Second, the two point sets (in projector and camera coordinates) are undistorted, i.e. the lens distortion is removed with the intrinsic parameters, and rectified. Third, the rays from the optical centres and the respective image coordinates of the undistorted points are transformed to the world coordinate system, which is the camera coordinate system. Finally, the intersections between the corresponding rays can be calculated yielding 3D points in world coordinates [90, 104]. This point cloud is manually filtered, resampled, and cleared of outliers. From the filtered point cloud a surface mesh is generated with the algorithm by Marton et al. [216].

The acquired data then needs to be processed to render information on the surface of the operating field. First, the projection surface point cloud is used to determine which projector pixels correspond to the world coordinate of the surface. With the help of the projection matrix of the projector derived from the calibration step, each point is projected to its two-dimensional position in pixel coordinates. Not every pixel position is represented by a known world point because of the low density of

Figure 4.13: Projection of a planned insertion point and contours of a rib on the irregular surface of an abdomen phantom viewed from the feet side in the correct viewing position. ©2018 John Wiley and Sons

the point cloud compared to the dense pixel structure. The gaps are filled through linear interpolation.

Next, a ray casting is performed to determine which structures are visible from a certain point of view and to determine where to draw them. For each projection pixel, a ray originating from the user's fixed point of view and pointing towards the pixel's respective world coordinate is checked for intersection with a world object. Because a fixed viewing position is assumed, some static structures only need to be processed once while other dynamic structures are updated frequently. The calculated depth values for all objects are then combined to generate the projection image.

The visualisation may include risk structures, the target for the intervention, a virtual part of the needle that is inside the body, and a virtual extension of that needle. Risk structures, such as blood vessels, and other anatomical parts, e.g. ribs, are optional and rendered transparently to both indicate depth and to avoid occlusion of the target (see Figure 4.13).

4.3 Needle Guidance

To support needle guidance within the interventional MRI, two visualisation concepts for navigation cues were developed on the basis of the visualisation techniques suitable for AR described in Section 3.5. Following the workflow of

MRI-guided percutaneous tumour ablations in Section 2.2.1 and the proposal of Seitel et al. [191], the needle insertion procedure is split into three steps:

1. Positioning the needle tip at the planned entry point.

2. Aligning the needle to the desired orientation to match the planned path to the target.

3. Inserting the needle to the target while maintaining the correct orientation.

During these steps the needle tracking yields the pose of the needle. In each frame, the needle's position and orientation are compared to the planned path. The needle is virtually extended along the shaft axis in order to extrapolate its intersection point with the surface tringle mesh of the operating field, which was created from the point cloud acquired with a structured-light scan. This intersection point is then compared to the planned entry point, of which the visualisation is adapted accordingly. As a measure for the orientation, the angle between the needle's direction vector and the vector between insertion and target point is calculated. This data is then visualised depending on which visualisation concept is used.

4.3.1 Navigation by Explicit Aids (*2D*)

The concept *2D* uses explicit aids to guide the user through the single insertion tasks positioning, alignment and insertion as illustrated in Figure 4.14. The positioning of the needle at the planned insertion point is supported by a coloured circle of adaptive size. The nearer the user sets the needle intersection point to the specified position, the smaller the radius of the circle gets. This is further elucidated by a colour gradient from red (high distance) to green (small distance).

After placing the needle tip at the planned entry point, the needle shaft needs to be aligned to point to the target. This process is supported by displaying an arrow originating at the insertion point and pointing towards the direction in which the needle shaft needs to be tilted. The arrow orientation depends at any time on the current deviation of the needle shaft to the target, so that the target can be hit even when the entry point was not targeted accurately. The length of the arrow results from the angle between the needle direction and the designated path. The smaller the angle gets, the shorter the arrow is drawn. The length is logarithmically interpolated so that alignment changes at smaller differences to the planned orientation have less effect on the arrow length than changes at larger angle differences. That way, the arrow remains visible at small alignment deviations and thus allows for fine adjustments. At an angular difference below 1° the arrow's colour changes from red to yellow indicating an acceptable alignment. Further

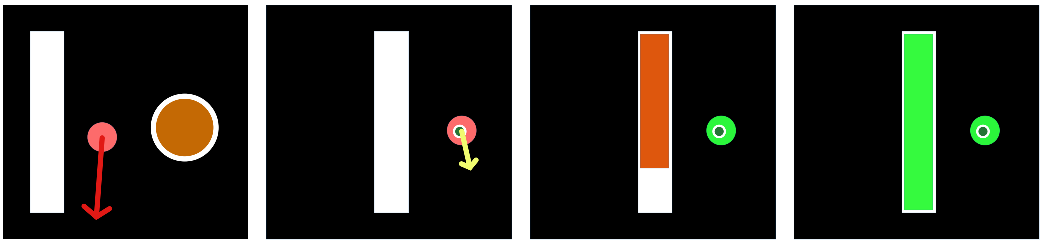


Figure 4.14: Visualisation of two-dimensional explicit navigation aids while moving the needle to the entry point, aligning, and inserting it. The needle-surface intersection is shown as a red dot, the planned insertion point as an orange/green circle with white borders (distance dependent), the desired orientation as a red and yellow arrow depending on the angle difference and a depth progress bar. From left to right: The needle is positioned next to the planned insertion point; the needle is set onto the planned insertion point, which is now small and green, and almost correctly aligned ($< 1°$, arrow is yellow); the needle is positioned and aligned as planned (arrow disappeared, insertion point is green), and has partly been inserted into the body; target is hit accurately (distance<0.2 mm) (depth bar green, insertion point green).

reducing the deviation below 0.5° causes the arrow to be replaced by a green dot, thus signalling a successful needle alignment.

Finally, the needle is inserted into the target. The depth of the needle is explicitly highlighted by the visualisation of a progress bar. The filling of the bar is linearly dependent on the Euclidean distance between needle tip and target point. The bar begins filling up after inserting the needle. Reaching a distance below 0.4 mm, the bar's colour turns from red to yellow. When further reducing the distance to 0.2 mm, the progress bar and the needle-surface intersection point change their colour to green, thus implying a successful needle insertion. Over-inserting causes the colours to change back to red and continues downwards from the progress bar.

Due to the flat visualisation of explicit aids, this concept is referred to as concept *2D* in the following. The colours where chosen strong and opaque in order to level out possible transparency effects occurring when projecting images (see Sections 3.2.4 and 3.5.2).

4.3.2 Navigation by See-Through Vision (*3D*)

Concept *3D* enables the user to virtually see through the projection surface by visualising segmented structures, the aimed target, and the needle inside the body.

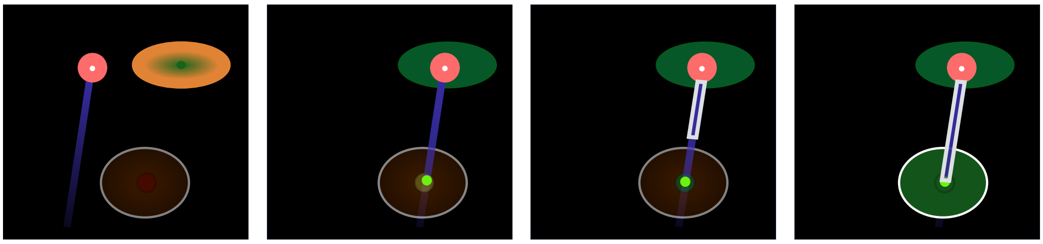


Figure 4.15: Navigation visualisation to support needle navigation. To indicate depth, deeper structures are rendered transparent according to their depth from the surface. The needle-surface intersection point is rendered as a red circle and the needle is extended with a blue line. From left to right: The needle is placed next to the planned entry point (orange green circle); the needle is at the entry point (now completely green) and oriented approximately correctly (green dot shows intersection at target – target centre is yellow); needle is inserted partly (virtual silhouette of real needle is drawn to indicate depth) and the needle is perfectly aligned (inner target sphere is green); the target was hit (everything is green) ©2018 IEEE

The concept is illustrated in Figure 4.15. These visualisations are rendered with the help of ray casting and show structures perspectively correctly from the fixed viewing position. For each pixel, a ray originating from the viewing position is cast through the scene using the world coordinate correspondences. These rays are then checked for intersections with the surface triangle mesh and the cylinder representing the needle and its path. The resulting depth values at the intersections are then compared with the originally calculated depth maps of the static objects. The depth values are then sorted and a pixel colour is calculated taking into account the colour and transparency values of the respective intersected objects. The resulting colour of an object is the product of a specified base colour, an applied Phong shading coefficient, and a depth encoding coefficient. The latter is a linearly interpolated factor that depends on the distance to the viewing position. The greater the distance, the smaller the coefficient and the darker the final colour. Because the final image is displayed with a projector, the darkening of colours results in a transparency effect as explained in Section 3.5.

In order to support the user with this concept, the needle is represented as a virtually elongated cylinder. In this way, the user can see the path the needle would follow if it was inserted at that moment. A coloured dot along the path line indicates where the virtual needle extension intersects an object. The colour of this point depends on the type of object being intersected. Red indicates the intersection of risk structures, while green indicates the intersection of the target structure.

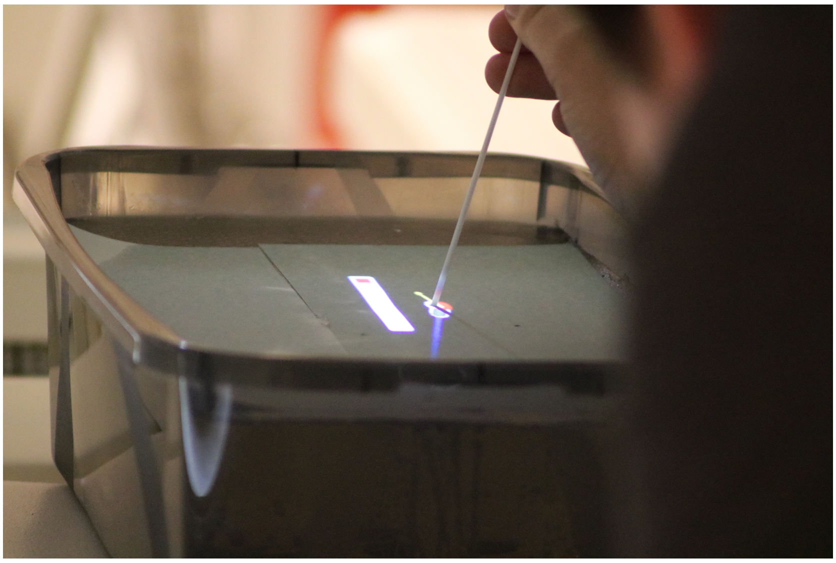


Figure 4.16: Exemplary view over the user's shoulder of the AR needle navigation inside the MRI bore. A combination of both *2D* and *3D* concept is shown. ©2018 IEEE

This concept uses a different approach than concept *2D* to visualise the planned entry point. The point is represented by a circle on the projection surface with a constant radius. The border of the circle is red, while its centre is coloured in green. The space between these colours is interpolated. The closer the needle is to the insertion point, the greater the green colour of the centre. A perfectly positioned needle produces a completely green circle. To make this even clearer, the colour lights up when the needle is correctly positioned.

In addition, the target area is displayed as a two-zone sphere. In this way, the target is shown large enough to be noticed quickly, while the inner zone emphasises the centre of the target more precisely. The two zones are also used for colour coding. The colour of the inner zone represents the state of the planned alignment of the needle and corresponds to the previously described colours of the alignment arrow of concept *2D* in Section 4.3.1. The colour of the outer zone represents the Euclidean distance from the needle tip to the centre of the target and is coloured similarly to the progress bar of concept *2D*. A perfectly aligned needle in the centre of the target results in a target ball with two green zones.

This concept makes use of the visualisation of three-dimensional structures inside the body. Therefore, this concept is called concept *3D* in this work. A real world view of the needle navigation visualisation on a phantom is shown in Figure 4.16. For demonstration purposes, the image shows both visualisation concepts at the same time.

4.4 Evaluation

A user study was carried out to assess the accuracy and usefulness of the navigation concepts. Eight participants took part, four of whom were radiologists with 4 to 15 years experience with minimally-invasive needle-based image-guided interventions and four users with technical background in the medical field. The users were given the task to insert a tracked needle into a phantom body inside the MRI bore without image guidance, relying only on the provided navigation concepts projected onto the operating field.

4.4.1 Apparatus

Phantoms were built of a plastic box filled with candle wax and rubber o-rings with a diameter of 2.3 cm serving as targets. The candle wax is clearly visible in T1 sequence, introduces no significant noise to the imaging, and provides good contrast to the rings. The phantom was covered with thin cardboard so that the targets and the needle could not be seen by the users and placed at the isocentre of the MRI.

The thresholds for the needle guidance colour indicators were set very close to the target, so that as accurate results as possible could be achieved. The distance threshold to indicate a perfect target hit was set to 0.5 mm, the angle difference between needle and target was 0.5°.

4.4.2 Study Procedure

The needle had to be positioned on the pre-planned entry point, aligned as required and then inserted until it hit the target. Each user was given unlimited training time before each concept until they felt confident with the navigation system. The task was repeated by each user three times for each concept individually with different entry points and targets. The order of the visualisation concepts for each user was varied between the subjects using the biased coin method: the two concepts were assigned a side of a coin each which was flipped. The concept assigned to the upper side was to be used first by the user, the other second. The next user performed the tasks with the reverse order of concepts. The coin was flipped for every second user.

After each insertion, an MRI data set of the phantom and the needle was acquired. In the end, an expert interview was carried out with each participant. The questions included valuations of advantages and drawbacks of the AR system and its suitability for the task.

Figure 4.17: View of the projected scene from above with the wide-angle camera. ©2018 John Wiley and Sons

4.4.3 Measures

First, the calibration accuracy indicators were determined. The reprojection errors of camera, projector, and stereo calibration from five different calibration procedures were calculated. Further, the standard deviation of the pose estimation of the ChArUco markers was calculated for 100 measurements to evaluate repetition accuracy.

The puncture duration, as well as errors between the planned and reached entry and target positions were measured in the post-interventionally captured data set with the integrated distortion correction of the MRI active in order to best maintain the validity of the 3D positions. After each concept, each user answered the meCUE [217] (for assessing user experience) and Nasa RawTLX [218] (subjective workload) questionnaires. The latter is an adaption of the original NasaTLX questionnaire [219], but without weighting the different scales. The answers give insights into usefulness, usability, intention to use, and subjective workload.

4.5 Results

The mean camera calibration reprojection error was 0.40 ± 0.19 px, that of the projector 0.62 ± 0.28 px, and the stereo calibration resulted in an reprojection error of 0.94 ± 0.33 px. The repeated 100 pose estimations of the ChArUco markers used for the calibration revealed high repeatability with a standard deviation of 0.1 mm in x direction, 0.2 mm in y direction and 0.7 mm in z direction at a distance of 650 mm, which is the distance between wide angle camera and the marker fixed to the MRI. The Rodrigues angle axis only differed by (0.003;0.001;0.004) and a standard angle deviation of 0.003°.

Table 4.1: Results of the accuracy and duration measurements grouped by experience and concept.

experience	concept	duration [s]	entry point error [mm]	target distance error [mm]
med	*2D*	127.8 ± 45.4	2.1 ± 0.9	1.7 ± 0.5
	3D	96.60 ± 41.2	1.9 ± 1	1.5 ± 0.4
tech	*2D*	85.25 ± 30.9	1.9 ± 0.4	2.3 ± 0.5
	3D	106.5 ± 33	1.4 ± 0.3	1.8 ± 0.6
all	*2D*	106.6 ± 43.8	2.0 ± 0.7	2.0 ± 0.6
	3D	101.6 ± 36.9	1.7 ± 0.8	1.7 ± 0.5

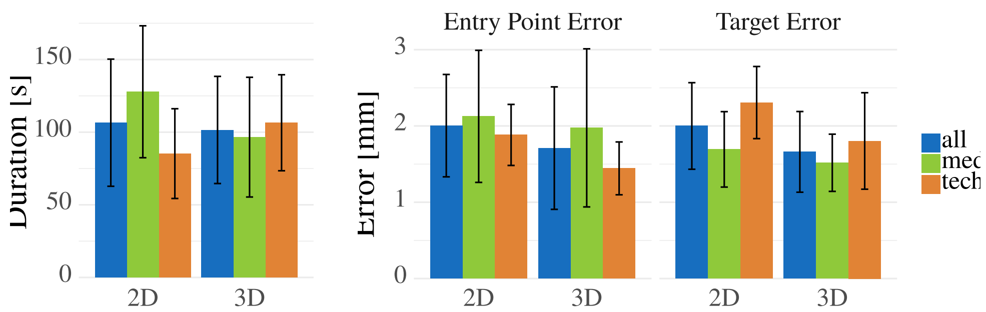


Figure 4.18: Measured task duration, entry point error and target distance error for each concept grouped by experience (all, medical, technical).

All users successfully finished training the needle puncture with the AR system after one test run and completed all tasks with both concepts successfully. However, the study did not reveal substantial differences between the concepts regarding positioning errors or puncture duration between the groups or concepts. The results are presented in Table 4.1 and Figure 4.18. The medical users were slightly faster with *3D*, but slower with *2D*. However, they were not as accurate as the technical users both in finding the entry point and the target with both concepts, but with neglectable difference. The overall target error of both concepts is very low with 2.0 ± 0.6 mm (*2D*) and 1.7 ± 0.5 mm (*3D*). The entry point was found with similar accuracy (2.0 ± 0.7 mm (*2D*) and 1.7 ± 0.8 mm (*3D*)).

The subjective workload scored slightly lower for concept *2D* (33.5 ± 12.8) than for concept *3D* (38.1 ± 8.3), but still similarly low, and nearly the same for both groups (see Figure 4.19). However, frustration is much higher with *3D* (65.6 ± 9.8) than with *2D* (35 ± 196), but mental demand was higher with *2D* (37.5 ± 12.2) than with *3D* (25.6 ± 16.1). The technical users reported a higher

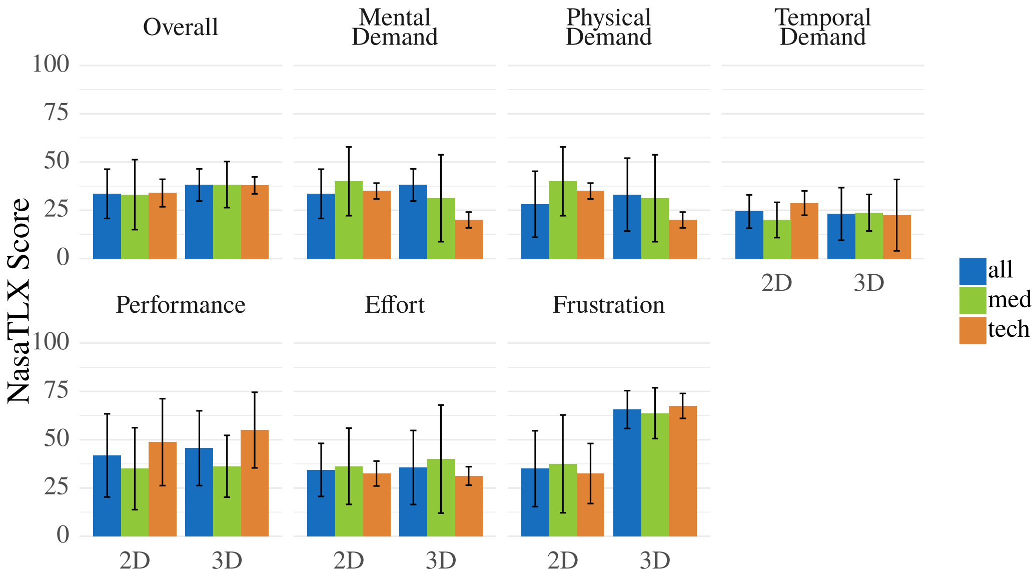


Figure 4.19: Nasa TLX results for each concept.

perceived performance in both concepts (*2D*: 48.8 ± 22.5; *3D*: 55.0 ± 19.6) than the medical users (*2D*: 35.0 ± 21.2; *3D*: 36.3 ± 1.6).

The meCUE scores for both concepts are shown in Table 4.2. All users rated the usability of both concepts high, the usefulness was valued higher among the technical users as well as the intention of use. The overall user experience rating shows that the users favoured *3D* over *2D*.

The users' valuations of the AR navigation from the interviews were predominantly positive. All users agreed that the navigation clues in general helped them guide the needle from the insertion point to the target in a clear and reliable way. All users experienced drops in the needle tracking speed while orienting the needle at the entry point before the insertion, caused by the distance to the MPT camera that was too short. This was perceived as a noticeable disturbance. Full tracking speed was only reached from a distance of 25 cm from the camera. Some users of both groups noticed a partial covering of the projection with the hand while guiding the needle, especially with concept *3D*, which could only be avoided by repositioning the hand.

The radiologists assessed the AR navigation system as suitable for needle punctures. When asked whether breathing motions require compensation, two of them stated that this is not a mandatory feature, because the insertion could be done while the breath is held, for which valid planning data already exists. They also acknowledged that dynamic registration and planning updates are complex and subject of extensive research, which was not feasible at the moment of the study.

Table 4.2: meCUE scores for both visualisation concepts. All scales are measured on a 7 point Likert scale, except the overall rating, which is from negative 5 to positive 5.

scale		usability	usefulness	intention of use	overall
med	2D	5.6 ± 1.3	4.5 ± 0.7	3.7 ± 1.0	1.9 ± 0.9
	3D	5.8 ± 1.4	4.9 ± 0.3	3.9 ± 1.1	2.8 ± 1.3
tech	2D	5.6 ± 0.6	5.4 ± 0.7	4.3 ± 0.7	1.9 ± 0.3
	3D	5.2 ± 0.4	5.2 ± 0.8	4.0 ± 0.3	2.9 ± 0.5
all	2D	5.6 ± 1.0	5.0 ± 0.8	4.0 ± 0.9	1.9 ± 0.6
	3D	5.5 ± 1.0	5.0 ± 0.6	4.0 ± 0.7	2.8 ± 0.9

All medical participants agreed that the AR navigation system is a valuable support for MRI-guided needle interventions, because the guidance allows for correctly oriented needle insertion for the important first stage during the insertion. When oriented correctly from the beginning, fewer corrections are needed in deeper sections, which would injure more healthy tissue or possibly move the target. None of the participants would rely solely on AR navigation, because real-time imaging is essential to ensure the distance to risk structures at all times. However, this would be displayed on the in-room screen of the MRI.

4.6 Discussion

As the results show, the proposed AR system is suitable for interventional needle guidance on a phantom inside a wide-bore MRI. The reprojection errors of the camera, projector, and stereo calibration are below 1 px, which is reported to be appropriate to achieve accurate results [104]. The standard deviation of the ChArUco marker fixed to the MRI is infinitesimally small. Therefore, these markers are perfectly suitable for optical pose estimation inside the bore, taking into account the MRI safety requirements.

The quantitative measurements of planned and actual needle insertion points and target points, as well as their errors, reflect the accuracy of the calibration and registration process. Compared to the accuracy of the interventions guided by MRI real-time imaging proposed by Rothgang et al. [13] (1.8 ± 1.5 mm) and Meyer et al. [67] (4.0 ± 1.2 mm), the needle guidance method introduced by Busse et al. [21] (2.2 ± 0.6 mm), the robotic assistance by Moche et al. [24] (2.2 ± 0.7 mm) and the HMD-based AR system used by Wacker et al. [203] (1.1 ± 0.5 mm) the target positioning errors of the AR system proposed in this work (*2D*: 2.0 ± 0.6 mm;*3D*:

1.7 ± 0.5 mm) are similar. Unfortunately, no entry point errors were reported in the literature. The mean puncture duration (*3D*: 106.6 ± 43.8 s; *3D*: 101.6 ± 36.9 s) is slower than a puncture under real-time imaging guidance (37 ± 14 s) [67], but depends on the circumstances of the intervention, i.e. the depth of the tumour inside the body and its accessibility, and thus cannot be compared directly.

Nevertheless, it must be noted that the measurements from MRI data sets are inherently faulty due to the relatively low imaging resolution and the virtual malformation of the scanned object when processing the retrieved signals. The latter leads to straight lines appearing as curves. As an attempt to best preserve the validity of the 3D positions in the data sets, the built-in interpolation and distortion correction algorithms of the MRI were activated. Because the test phantom was not being moved during the study, all MRI scans yielded best possible reproducible results. Another error source is the stiffness of the applied needle. As the needle progressed deeper into the test phantom, it was slightly deformed. Because the MP markers were placed at the top of the needle, the needle tip position was marginally different from the expected location estimated by the tracking system. This can be avoided by using a stiffer needle material or extracting the needle artefact from the real-time images and considering the needle curvature in the pose estimation [220].

The measured accuracy results reflect the users' positive impressions of the AR needle guidance system. The visualisation of the navigation aids and the virtual needle was perceived as perspective correct, although it was fixed at an assumed viewing position. Nevertheless, in the future, the user's viewing position could be tracked to ensure a correct perspective and use motion parallax to achieve a more immersive visual representation of 3D objects. This is assumed to only provide a small effect, because the viewing position is restricted to a small volume at the MRI's feet side due to the workflow and patient access limitations.

The low MPT rate caused frustration to the users. When the MPT markers were too close to the MPT camera, the tracking stopped and the users could not get feedback on the position or orientation of the needle. This resulted in more correcting movements, i.e. pulling the needle back and inserting it again, in both concepts. Additionally, according to the RawTLX results, concept *3D* shows a higher potential for frustration. Some users commented that they were confused when aiming for the target because of the inverted movement of the virtual needle extension compared to the shaft when the tip had not been inserted deeply, such that they were not able to immediately align the needle as planned. Thus, some users had to correct more punctures as compared with concept *2D*. However, the accuracy of both concepts and the usability and usefulness scores were similar, thus no superior concept could be determined in this respect.

Some users commented they were more comfortable with one concept than with the other, which underlines that the choice of concept is not a question of suitability, but more of the user's personal preferences. Furthermore, no notable difference between the two user groups could be found in the study regarding accuracy or subjective perception of workload or usability. This could be due to the easy-to-understand and unambiguous visualisations, which do not presume medical experience, as some users from both groups agreed. The clear position and alignment feedback during the whole process with colour and shape changes of the visual aids compensates for the differences in experience.

Nevertheless, by tracking the user's head position, further development steps could enable kinetic depth cues by interactively adapting the visualisation to the viewing position. Other illustrative visualisation approaches than silhouette enhancement could be used to encode more depth information on the needle's progress or distance to structures of interest [137], which were not subject of the presented study. These additional depth cues could improve user performance and reduce frustration by better clarifying the three-dimensional virtual scene. A combination of both navigation concepts should be considered especially when risk structures and the needle's distance to them shall be highlighted as proposed by Heinrich et al. [221].

Overall, the users confirmed with high meCUE usability and usefulness values as well as in comments that both proposed concepts for visualisation of needle guidance serve as suitable support for MRI-guided interventions. As can be seen from the intention of use score, the users would use the introduced projector-based AR navigation system beyond the scope of the study. The ergonomics of this setup were assessed as acceptable by the users, because they did not need to turn their heads to the external monitor next to the bore. This can prevent back and neck pain, but needs further long-term investigation.

Despite the promising accuracy assessment, the AR needle guidance has some inherent limitations when it comes to interventions in real clinical scenarios. As shown in Figure 2.3a, a thick flex coil is placed on the patient that is needed for imaging but covering the actual operating field. This coil should be integrated into the sterile drape so that the projection can reach the patient's skin (see Figure 4.20) [222]. Because the needle guidance relies on non-real-time planning data, the needle can only be inserted when the planning data matches with the operating field. This can be reached by arresting the patient's respiration. The timing must then meet the same position as when the planning data was acquired. Although, this is not always possible. Liver tumour ablation, for example, is mostly performed after completely immobilising the patient anyway, therefore the breathing could be halted by the ventilator during the insertion process. For biopsies, the patient is mostly not anaesthetised, therefore the breathing cannot be accurately controlled. This matter must be further evaluated to allow for a sound estimation of the breathing movement. However, there are a few solutions to keep track of organ movement involving dynamic registration of live 2D MR images on a pre-interventionally acquired high resolution data set [223] or generating a

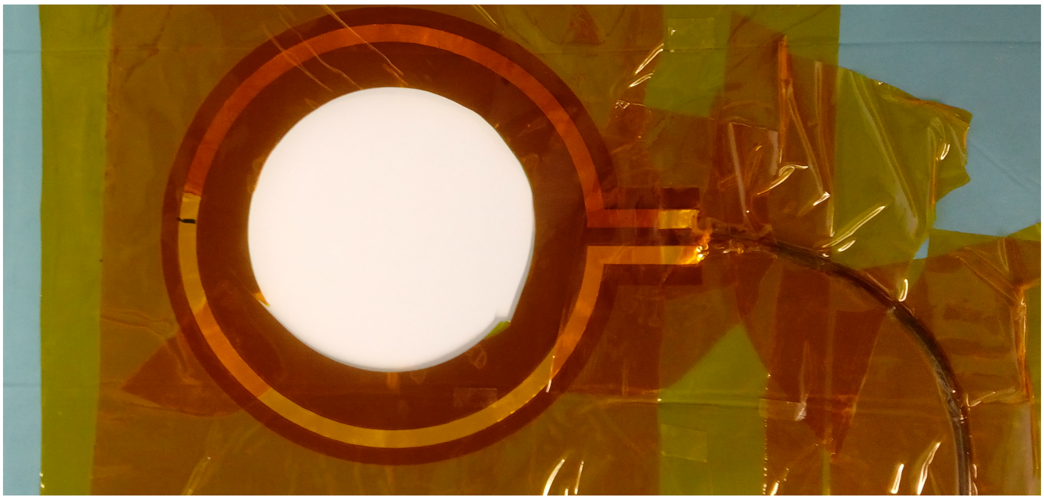


Figure 4.20: A flexible MRI coil integrated in a sterile drape for interventional use [222].

predictive breathing model from live MRI registered to pre-interventional image data [224]. These should be investigated in the future, too, in order to update the projected navigation information. To react to changes of the projection surface due to breathing movements, especially for abdominal treatments, an approach to dynamically map the projection should be followed in further research, e.g. by using markers on the surface to track deformations [225]. It may be possible to integrate markers into the incision drape that is placed on the operating field in advance of the intervention. However, this depends heavily on the patient positioning, because an entry point on the lower side of a patient could neither be reached by the projection, nor be seen by the camera. In this case, the hardware must be set up differently. More important than skin movement, breathing also affects organs and surrounding risk structures through deformation. With additional real-time control imaging, this problem could be addressed. In further research, real-time images of the needle plane could be colour-corrected and overlaid on the patient to give the radiologist the certainty of not damaging risk structures due to outdated image data.

According to the feedback from the clinical users, the AR system is a valuable support for needle guidance, especially during the important first stage of the insertion where the needle needs to be as accurately oriented as possible to prevent elaborate corrections later on that would hurt more healthy tissue than necessary. The projection interface should also be considered as an aid to convey the movements of the needle under the skin to improve the coordination of inexperienced radiologists.

4.7 Conclusion

In this chapter, the first projector-based AR system for the use inside an MRI bore has been presented. It was successfully set up in a way that considers both the workflow of MRI-guided tumour ablations as well as the restrictive environment the MRI introduces. Complying with the workflow, the AR system does not limit the MRI to a mere offline image provider but enables physicians to also benefit from morphological and functional imaging, though on a separate screen. The system was constructed with a long-throw projector and a wide angle camera together with an optical MP marker needle tracking system. The components were calibrated with respect to the world coordinate system as well as the MRI and patient coordinate systems. A fixed optical marker at the bottom of the bore provides persistence of the calibration. The AR system was used to display a clear visualisation for needle guidance that is easy to understand and accurate.

Two concepts for the visualisation of needle guidance aids aligned with the operating field using the AR system have been presented. In the *3D* concept, a virtual depth-coded needle, the planned entry point and the target are provided. The virtual needle is extended to facilitate correct alignment with respect to the target. Colour changes reflect the current state and accuracy of the puncture. The *2D* concept does not include any 3D visualisations. Instead, the planned entry point, an orientation arrow and a progress bar are visualised with the current distance to the target. A user study with four experienced radiologists and four participants with a technical background in the medical field did not reveal a clear superior concept. In addition, only slight differences in accuracy or usability were found without favouring one of the concepts. The high meCUE and low NasaTLX scores confirm the ease of use of the visualisations with clear feedback that can be understood without prior experience with the AR needle guidance system. The accuracy of the AR navigation system, including measurement, user, and calibration errors, is comparable to the results of the related literature.

This prototype serves as a platform for further development, e.g. for AR visualisation techniques that improve depth perception during interventions, as proposed in Bork et al. [214], Hansen et al. [136], Lawonn et al. [137], or Heinrich et al. [221]. In addition, 1D MRI real-time images of the needle plane could be aligned with and projected onto the patient to enable uncomplicated and ergonomic access to information on the current needle pose and thus to overcome the problems introduced by breathing motion. To extend the support for risk visualisation, an error cone as proposed in Alpers et al. [226] could be added to the scene or coloured lines as a measure for the shortest distance between the needle tip and risk structures [221]

Additionally, for clinical applicability several technical restrictions need to be overcome. The size of the bore needs to be increased by the MRI manufacturers, otherwise, the integration of needle tracking is cumbersome and limited to a small tracking volume. Furthermore, it would be beneficial to integrate one or more projectors into the bore to avoid the steep projection angle that may cause the image to be occluded by the user's hand. This would as well improve the projector calibration quality. If applied as suggested, the proposed system has the potential to facilitate needle-based interventions inside closed-bore MRI scanners.

This chapter contains research that has been published in:

A. Mewes, F. Heinrich, U. Kägebein, B. Hensen, F. Wacker, C. Hansen. "Projector-based augmented reality system for interventional visualization inside MRI scanners". *The International Journal of Medical Robotics and Computer Assisted Surgery*, 15.1 (2019), e1950. [227]
A. Mewes, F. Heinrich, B. Hensen, F. Wacker, K. Lawonn, C. Hansen. "Concepts for augmented reality visualisation to support needle guidance inside the MRI". *Healthcare technology letters*, 5.5 (2018), pp. 172-176. [228]

The navigation concepts were first implemented by Florian Heinrich within his master's thesis, which was supervised by the author of this thesis.

5

Interventional Touchless Gesture Interaction

During image-guided procedures the radiologist performs the intervention relying on either live fluoroscopy images or a planning data set of the patient's anatomy in order to access lesions. The direct interaction with these is challenging in a sterile environment when using conventional input devices, such as mouse, trackball, joystick, or touchscreen, for several reasons. Input modalities that must be touched are prone to contamination with bacteria [229, 230] and must be covered with sterile drape to prevent infections if operated by the sterile hands of the radiologist hence affecting the usability of the input device. This often leads to the physician delegating the interaction to an assistant due to poor usability. This indirect interaction is time-consuming, error-prone [13, 19, 231] and requires additional specialised personnel, which, e.g., causes higher treatment costs. In an environment with loud noise, such as an MRI, communication with the assistant is difficult while sequences are run. Hence, the sequence must be stopped to be able to convey the commands correctly. Sometimes, the radiologist even leaves the sterile area to access additional patient data on a workstation, e.g. to measure distances between vessels or compare different data sets [13, 19, 231]. The controls for the imaging modality, such as a joystick or foot pedal, may also sometimes be out of reach or blocked by another user and must either be used by an assistant or the intervening radiologist must change the position to access them. This workflow is too elaborate and delays the intervention inappropriately [13].

To overcome the complexity of current user interfaces available for image-guided interventions and to facilitate the workflow, a sterile, touchless, and metaphoric hand-gesture-based NUI is presented in this chapter which is used in two scenarios.

In a first scenario, a metaphoric hand and finger gesture set is presented that is used to control a medical image viewer during angiographic CT-guided interventions. The gesture set was evaluated regarding usability and then improved based on the user study results. As a follow-up, the improved gestures were compared against touchscreen interaction in a similar scenario in another user study [179].

In a second scenario, the touchless gesture control of an MRI scanner during MRI-guided interventions is introduced to provide a direct and sterile alternative to conventional input or delegating the interaction to an assistant. The gestures developed for the first use case were used to create a gesture set to control the MRI as well as to interact with the acquired images shown on the GUI displayed on the in-room screen. The proposed system is then evaluated in comparison with task delegation as the state of the art in clinical practice.

5.1 Gesture-Controlled Interactive Radiation Shield

In this section, a touchless gesture-controlled physician-computer interface is presented. It enables the physician to directly and sterilely interact with medical images at any time. It is integrated into the condensed workspace of an angiographic CT suite by projecting the medical image viewer interface on a radiation shield, which has no other purpose than protecting the radiologist from X-ray exposure. Thus, it is always in reach and available to the physician. In order to maximise the usability, a metaphoric hand gesture set has been developed based on the guidelines for designing NUIs (see Section 3.6) that is used to rotate anatomical 3D models, slice through 2D images, window, select objects, zoom in and out, and pan the images. In the following, the hardware and software setup are presented as well as the natural gesture set that constitutes the main contribution. Then, the usability study is described and its results are discussed.

5.1.1 Related Work on Interventional Touchless Interaction

Touchless HCI with interventional software and devices has been approached in a variety of ways over the years. A thorough overview and discussion on them is given in a systematic review by the author of this thesis [146].

Voice-controlled laparoscopic light [232], camera [233], endoscope holding [234], robotic assistance [235], or anaesthesia [236], eye-tracking-based [149–151, 237] or hand-gesture-based [238] telerobotic assistance, control of OR lights [239], and intraoperative registration [240], as well as head-tracking-based control of endoscopic camera movement [241, 242] have been presented. Multimodal approaches employing eye-tracking and foot-interaction [152, 243] are promising [153] when direct interaction with interventional software is desired but the hands are occupied.

While these are mostly operated without hands, the interventional control of medical image viewers is mainly focused on hand gestures in order to accurately convey the intended actions to the software. Wachs et al. [154] presented a camera-based hand gesture control for the exploration of medical images. Though, this system offers only a small set of gestures and lacks accuracy and robustness. Body-worn sensors are used [183–185] to recognise arm and hand gestures as input to be mapped onto a medical image viewer. However, these are only able to detect coarse movements and are not suitable for fine interaction with medical images if the interaction shall be fast. Using myo-electric signals of the forearm to detect finger gestures produces a high false-positive rate, but enables the user to interact with software without being restricted to a small interaction area [186].

The introduction of affordable off-the-shelf consumer market interaction devices facilitated the development of touchless interaction concepts for interventional use. Ebert et al. [156] introduced a system to control a medical image viewer by simple hand gestures using the Microsoft Kinect structured-light sensor. They propose a training of 10 min to get familiar with their system. Due to a recommended working distance of 1.2 m a large display is required for many tasks, which might cause space problems. Many similar systems based on the Kinect employing arm gestures have been proposed [155, 157, 159–162, 244]. Gallo [245] additionally implemented filters to reduce signal noise and improve accuracy, others used supplementary voice commands [158, 188] or the current context of interaction [164] to decrease ambiguity. As Wipfli et al. [167] conclude from the comparison study between mouse input, task delegation, and Kinect interaction, the latter is slower than mouse interaction mainly because of its inaccurate, large movements that are necessary to distinguish gestures unambiguously. Therefore, finer, more precise gestures are needed to accelerate the interaction, reduce fatigue, and enable the interaction in space-limited environments, such as medical intervention rooms.

With the introduction of the LMC it became possible to easily design hand and finger gestures. The introduction of this device led to an increased interest in interventional HCI research, similar to the effects the Microsoft Kinect entailed [146]. The LMC is proven to be accurate. Weichert et al. [246] performed accuracy measurements by determining the deviation between the desired 3D position of a reference pen and the 3D position obtained by an LMC. In static test cases, the deviation was lower than 0.2 mm, in dynamic cases it was below 0.7 mm. The diameter of the reference pen used as tracked tool had no impact on the error. Similar results were found by Guna et al. [247], where the standard deviation was less than 0.5 mm in static test cases, but a significant drop in accuracy was recognised when samples were taken more than 250 mm above the LMC.

Bizzotto et al. [173] first implemented an LMC-based gesture control plugin for a medical image viewer triggering keyboard shortcuts with hand gestures. This is a common approach on gesture-based interaction and followed by many researchers [175–177, 181]. However, this is only an on-top solution on existing GUIs and does not allow for a natural integration of the gestures into a NUI (see Section 3.6.1), thus lacking usability and memorability. Some of the gestures merely consist of hand positioning and orientation in a defined volume [175]. These are, however, not suitable for handling complex interaction such as 3D rotation or selection. Mauser et al. emphasise the high potential for the use of an LMC in interventional situations, but also the need for a careful use case analysis and appropriate gesture selection. Because their gesture set included ambiguities, they postulate that a lock/unlock gesture is mandatory. Rosa et al. [176] report from clinical experience with the Microsoft Kinect as well as with the LMC that the LMC is less tiring due to smaller movements and it does not require a distance as large as the Kinect needs in order to work. Opromolla et al. [180] designed custom gestures with focus on usability and memorability. The gesture set consists of one- and two-handed gestures for windowing, zooming, selecting, panning, resetting, releasing, and calling help functions. They report a disadvantage in the duration of the interaction, but good memorability. However, two-handed interaction is not appropriate during an intervention, because the radiologist mostly performs the main task with at least one hand and manages the interaction with software with the other hand. Therefore, one-handed gestures would be more suitable.

The previous attempts to use hand and finger gestures to control a medical image viewer were successful, but still not leveraging the full potential of this gesture sensor. The gestures were either only simple hand positioning or different numbers of visible fingers, but not metaphoric real world gestures users naturally perform with their hands, or on-top solutions which are not integrated well into the application. During image-guided interventions, one-handed, unambiguous, and metaphoric gestures are needed to increase usability, memorability and reduce mental demand during the interaction. In the following sections, a gesture-controlled interactive display is presented and evaluated that addresses these issues.

5.1.2 System Setup

As an exemplary scenario for the implementation, C-arm-CT-guided interventions were chosen. In this case, the interventions are performed with an angiography suite, which is packed with devices and provides only limited space to move around or place additional hardware (see Figure 5.1).

The results of a workflow analysis conducted earlier [231] suggest that the additional patient data from the workstation outside the intervention room should be brought to the direct vicinity of the radiologist. In order to provide these image data an additional display is needed. However, the limited space restricts the introduction of an additional screen. To this extent, the CT's radiation shield, which protects the radiologist during the intervention, was equipped with a projector displaying a medical image viewer on its surface. This way, the image data is always in reach without the need to move around.

Figure 5.1: The angiography suite provides little space for additional displays. Therefore, the radiation shield on the left of the radiologist is used as a projection surface.

Hardware Components

A prototypical setup of the interactive radiation shield in the angiographic lab is shown in Figure 5.2. The ASUS P1-DLP pico projector with a resolution of 1200×800 *px* is placed behind the radiation shield to back-project the GUI onto a projection foil (Modulor Opera) attached to the shield. An LMC is placed below the shield such that all gestures are performed within the interaction volume in front of the GUI. This volume with a size of approximately $50 \times 50 \times 50$ *cm* is calibrated with the projection surface: the corners of the surface are pointed at one at a time and the finger tip coordinates are mapped accordingly onto the projector pixels. This allows for moving a cursor along the image viewer without scaling the hand movement.

Medical Image Viewer

To reduce complexity and provide the needed information during image-guided interventions, the software functions were chosen carefully depending on the use case of CT-guided interventions. From a discussion with the clinical partners, the following frequently used functions were determined:

- Navigation within a set of radiologic images including slicing, zooming, and panning of 2D images.

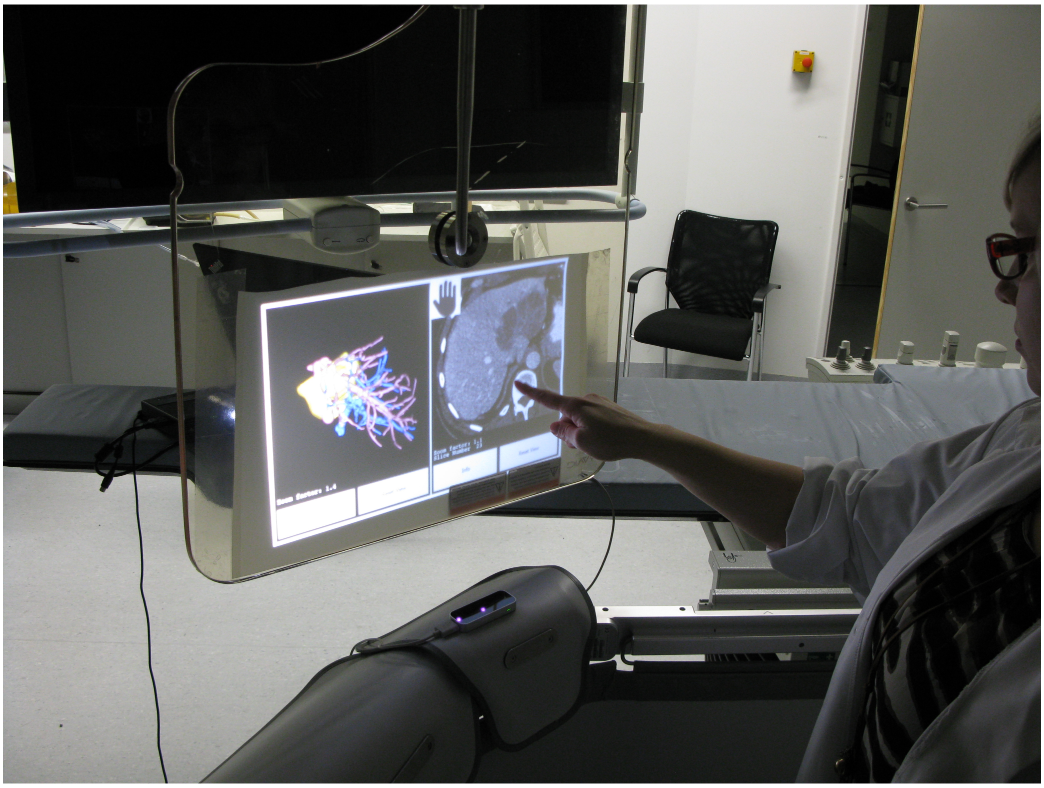


Figure 5.2: Prototype of interactive radiation shield. The GUI is back-projected onto the radiation shield. The gesture sensor is placed below. ©2015 CARS

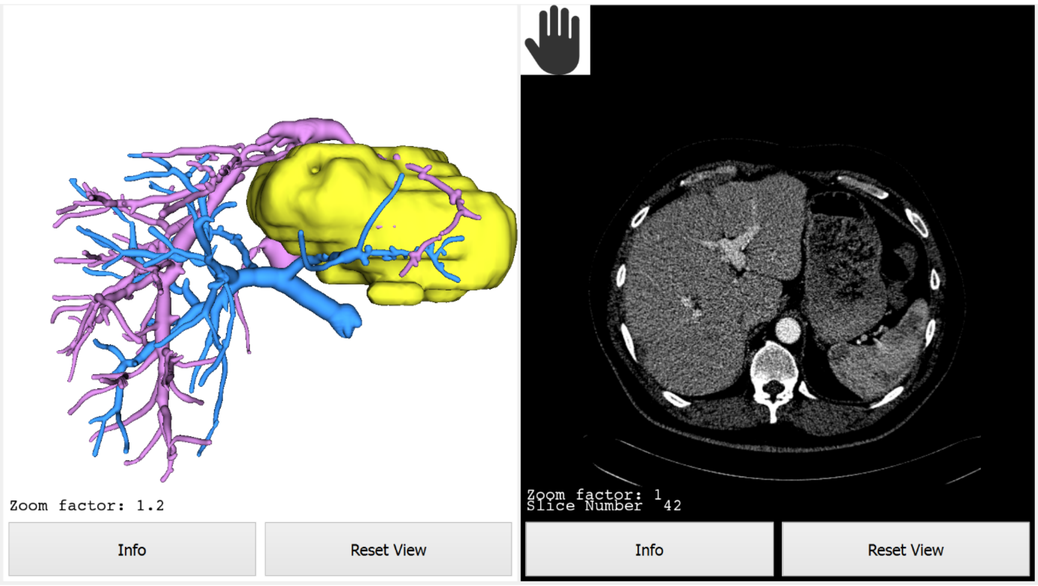


Figure 5.3: The GUI of the image viewer. It is split into two parts: a 3D model viewer on the left and a 2D image viewer on the right. The hand icon on the top indicates the currently recognised gesture of the right viewport. ©2015 CARS

- Manipulation of 3D planning models, i.e. rotation, zooming and panning.
- Selection of anatomical structures and triggering of other functions, e.g. via buttons.

Figure 5.3 shows the GUI of the projected image viewer. It consists of a 3D model viewer on the left and a 2D viewer on the right. Textual information is given on the lower left corner and buttons are provided for resetting the view and triggering another dummy function. Visual feedback on the currently detected gesture is provided on the top left corner of each viewport.

5.1.3 Gesture Set

The prototype was aimed at providing an NUI based on hand and finger gestures. In accordance with the guidelines for NUIs [142] and Nielsen et al. [248] these should be metaphoric, easy to memorise, self-describing, not stressing to the user, and logically connected to the software. Therefore, the proposed interaction set is inspired by interactions with real-world objects. This way, it is easy to remember and self-describing. To rotate a 3D object continuously, the user has to perform a *grab gesture* with a half-closed fist, as if grabbing it in reality, and rotate the

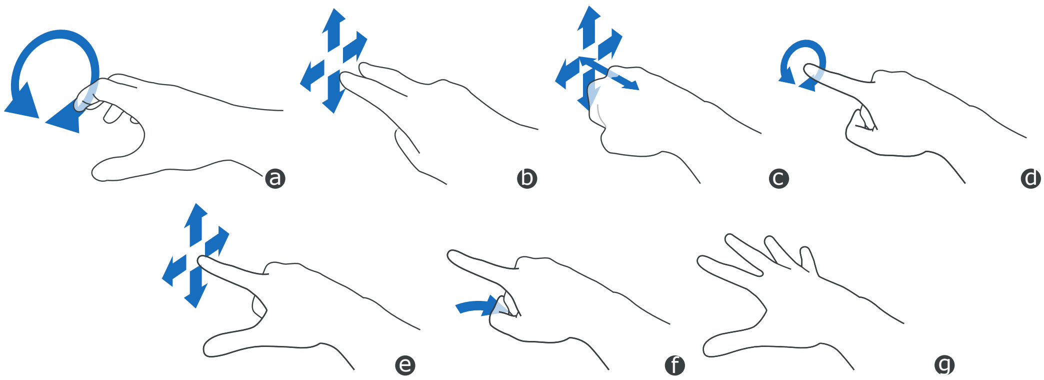


Figure 5.4: The gesture set. a) *grab*/rotation, b) *swipe*/rotation, c) *fist*/zooming/panning, d) *circle*/slicing, e) *pointing*/cursor movement, f) *pointing*/click, g) *open*/nothing.

hand (Figure 5.4a). A discrete 45° rotation can be achieved with a *swipe gesture* in up/down or left/right direction (Figure 5.4b). This is inspired by a globe rotation, but in two dimensions. Zooming is done similarly in the 3D and 2D view by virtually grabbing the 3D object or the slice, respectively, with a fully closed fist and pulling it from or pushing it towards the screen (Figure 5.4c). The user can change the slices by drawing circles with his index finger clockwise or anti-clockwise (Figure 5.4d). Slices can be moved vertically and horizontally with this gesture as well. This is analogous to turning/rotating a large control knob. The radiuses of the circles are mapped onto a step size while cycling through the images. For selection tasks, the user has to perform a *pointing gesture* (Figure 5.4e) with the index finger extended towards the object, and by hitting the non-extended middle finger with the thumb, a single click is triggered (Figure 5.4f). When extending the fingers to an *open* hand (Figure 5.4g), no action will be executed.

5.1.4 Evaluation

To evaluate the gesture set and its integration into the image viewer, a user study was conducted. It is separated in two parts, the pilot study and the main study.

Pilot Study

The pilot study was carried out with a usability expert and two medical students. The participants were asked to use the functions of the UI one at a time to achieve specific, predefined task goals mirroring the different gesture-controlled functions. They were asked to comment on the robustness, self-describability, and usability of the gestures using the Think Aloud method [249]. Despite predominantly positive feedback, the pilot study showed problems with the *grab gesture*. On the one hand, the participants found that the recognition of the gesture is not reliable enough, which leads to frustration. On the other hand, they noted that continuous rotation is not absolutely necessary. Therefore, in the main study, the swipe gesture for discrete 3D object rotation about fixed angle steps was introduced. Both gestures have been tested and are compared in the results section. The training time was not extraordinary long (<10 min), thus it was not investigated further.

Main Study

The main study followed a within-subjects design. Each user performed 8 tasks that reflected the functions of the medical image viewer and made use of all proposed gestures.

Participants

The subject pool for the main study consisted of 12 participants aged 21 to 41 years (M=26.4 years), eight were male and four female, nine were right-handed, three left-handed. Six participants had a technical background in medical technology, four were medical students, and two were radiologists. One radiologist had two, the other sixteen years of clinical experience. 42 % (5 out of 12) of the participants stated that they had many years of experience with medical software, 92 % (11 out of 12) with two-dimensional medical data sets, and 75 % (9 out of 12) with 3D models.

Tasks

The applicability of the gestures was evaluated in eight tasks modelled by scenarios from the workflow of CT-based interventions. These scenarios were developed together with clinical experts and based on observations and results of the previous workflow study [231]. The tasks were:

- Which slices does the tumour extend over (image shown)?
- Choose slice 68.
- Zoom into slice 42 to double its size.
- Enlarge and move layer 26 so that the tumour touches the edges of the screen! Then display additional information (dummy function).
- Rotate the 3D view with the grab gesture so that the following image is obtained (image shown).

- Enlarge the 3D representation so that the tumour is displayed 1.5 times as large (image shown).
- Rotate the 3D view with the grab gesture and enlarge it so that the following image appears and then lock the view (image shown).
- Rotate the 3D view with the wipe gesture so that the following image is displayed and then display additional information (image shown).

The tasks varied in the number of consecutive gestures that had to be remembered and performed to solve them and thus in difficulty.

Measures

In order to determine the self-describability of the gestures they were printed on a sheet of paper along with the software functions in a non-connected manner. The gestures had to be assigned to the correct functions after being demonstrated by the study supervisor. The number of false assignments yields an estimate of the self-describability of the gesture set.

A questionnaire was created (see Appendix) to collect qualitative information about the professional experience of users, in particular with gesture interaction or radiological images. In addition, two questionnaires with a five-point Likert scale were used. The first questionnaire lists gesture-specific questions inspired by [248] on the usability aspects mentioned in the requirements analysis. For each gesture, participants had the opportunity to leave a comment in the questionnaire.

The second questionnaire is based on the short version of the usability questionnaire ISONORM 9241/10 [250]. This questionnaire contains questions regarding aspects of usability, e.g. suitability for the task, self-describability, and error tolerance. The questionnaire was adjusted to be applicable to the used prototype. For example, the proposed UI has no features regarding individualisation and therefore, these questions were removed.

Procedure

The experiment was carried out with the proposed setup. First, the participants were described the setup of the gesture-controlled projection display. The different functions of the 2D and 3D viewer were explained. Second, the gestures were shown on a sheet of paper (similar to Figure 5.4) and the assigned function of each gesture was explained. Third, in the unlimited training phase, the subjects practised the gestures one after the other with advice from the supervisor until they felt comfortable using the software. After the training phase, which lasted a maximum of 10 min, the participants were asked to complete eight interaction tasks

Table 5.1: Durations of the gestures used in the subtasks of all tasks.

duration [s]	*circle*	*fist*	*grab*	*pointing*	*swipe*
Mean	32	37	92	6	72
Min	4	2	5	1	14
Max	160	146	245	52	145
SD	32	33	60	9	49

as illustrated in Figure 5.2. In order to avoid influences from learning effects, the tasks were randomised. This was also necessary because the tasks were of different difficulty. Finally, the questionnaires were filled out by the participants. In addition, an expert interview with the two radiologists was carried out to gain information about the applicability of the gesture interface in clinical practice. During the task execution the users were also asked to comment on anything notable using the Think Aloud method [249].

5.1.5 Results

The pilot study showed that the robustness of both 3D rotational gestures was insufficient for untrained users. Therefore, the gesture recognition was refined for the main study.

The gesture assignment tasks for estimating the self-describability of the gesture set led to an error of 2.6 on average (min: 0, max: 6). Almost all participants (10) assigned the *pointing gesture* correctly. Most mistakes were made with the *swipe gesture*. Here, only one participant assigned the gesture correctly.

The training times differed severely between 2 to 10 min. All participants completed the tasks successfully. The duration of each task was analysed to obtain the duration for each gesture execution. Because subtasks requiring the same gesture were designed to be equally difficult and time consuming, the gesture execution times shown in Table 5.1 indicate the difficulty of each gesture regarding robustness, accuracy, speed and include implications on suitability for the task, memorability, and other usability aspects. The standard deviation is very high for each gesture, especially with the grab or swipe gesture, i.e. for the 3D rotation tasks.

Figure 5.5 shows the results of the gesture-related questionnaire. For each gesture the users had to assess if the gesture was easy to execute, easy to memorise, natural, and not tiring. Three questions were specific to one gesture each: for the *grab* gesture it is relevant whether the way the hand is to be opened was comprehensible, for the *circle*, *fist*, and *swipe* gestures it was of interest whether the movements, e.g. the circle radius or the movement distances, mapped to the actions on the screen as expected. The results reveal weaknesses of both rotation gestures: both *swipe* and *grab* gesture are not easy to execute for 33 % of the users, the hand opening of the *grab* gesture was only comprehensible for 50 %. Further, the grab gesture was tiring for one third of the users. The *swipe* gesture was, however, rated natural by most participants (92 %). The other gestures were rather easy or easy to execute, felt rather natural, were rather not tiring, and rather easy or easy to memorise.

The results of the ISONORM 9241/10 questionnaire are shown in Figure 5.6. They reveal an insufficient error tolerance of the UI. Only 25 % of the users think that not much help is needed to learn the use of the UI. In total, the participants agree that the UI is suitable for the task and not many details need to be learnt to use it.

Table 5.2 summarises the comments of the users during and after the study.

Table 5.2: Pros and cons of the UI based on the comments of the users.

gestures	pros	cons
grab	simple and logical principle of grabbing is intuitive continuous rotation in real-time	tiring after a while not robust enough
swipe	similar to touchscreen, hence easy to learn and use	finger position is too important too little degrees of freedom fixed rotation restricts the user (should be adjustable)
fist	predictable easy to use	start and end point unclear
point	easy, simple, reliable	click was not very robust for some users
circle	precise and robust mapping mapping of the circling radius to step size is useful	

5.1.6 Discussion

The results reveal that the gesture interaction is suitable for the task and may integrate well into the workflow of CT-guided interventions. The gestures *circle*, *fist*, *pointing*, and *open* were rated appropriate for the interaction tasks and are not tiring, feel natural and are easy to perform. Problems were found with both 3D rotation gestures. The rotation was achieved faster with the *swipe* gesture than with the *grab* gesture, while the *grab* gesture was also rated worse in the gesture-specific questionnaire (see Figure 5.5). Further, the participants stated that even though

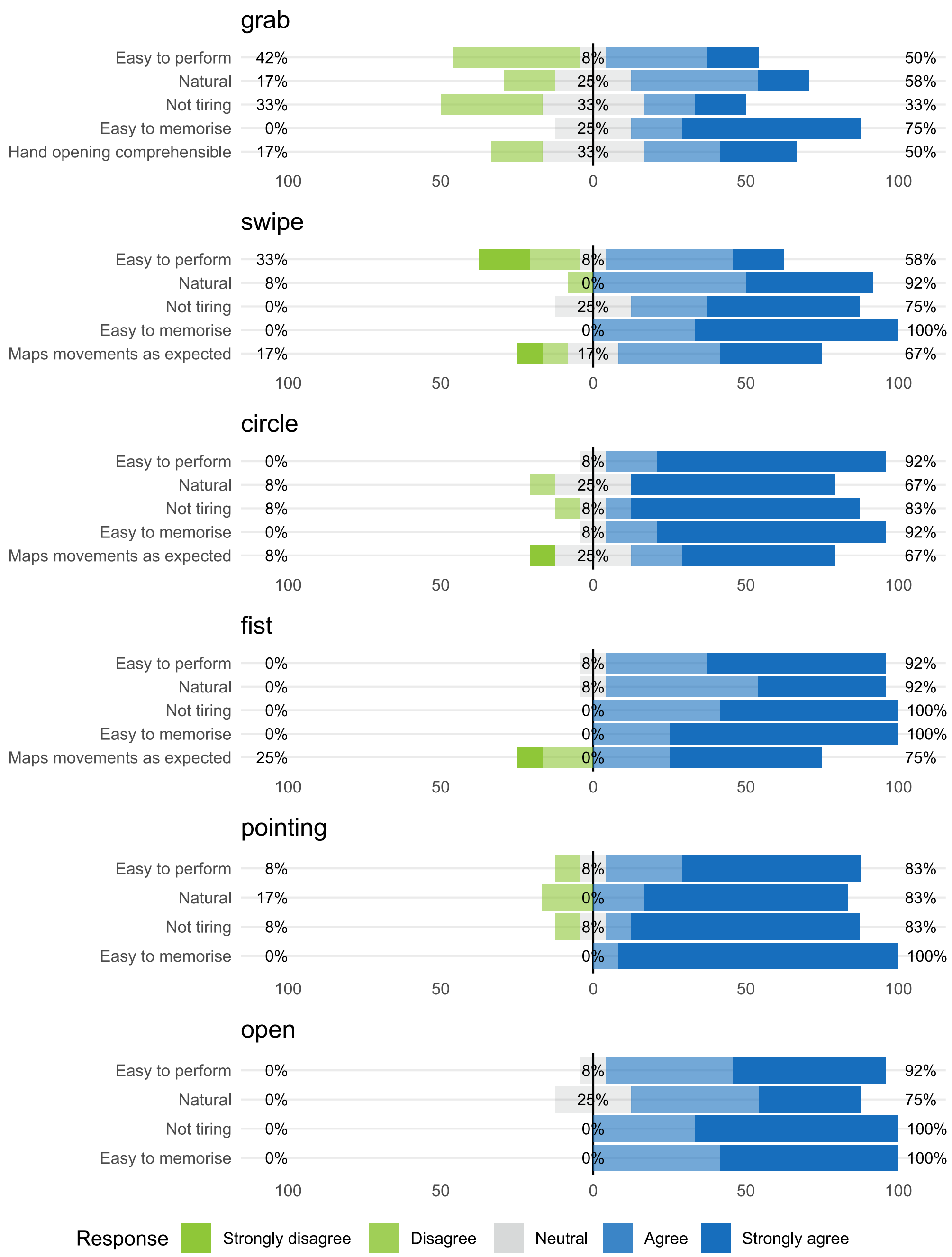


Figure 5.5: Users' valuations of the single gestures. "Hand opening comprehensible" was a question particularly for the *grab gesture*, "Maps movement as expected" was only asked for the *circle*, *fist*, and *swipe gesture*.

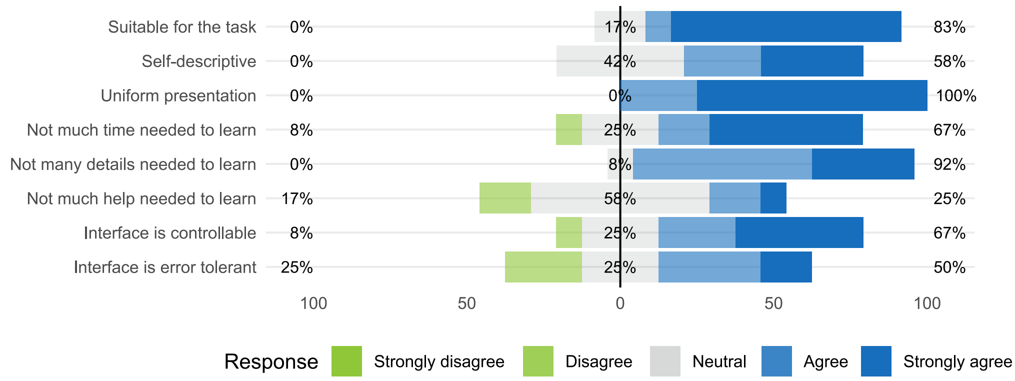


Figure 5.6: Usability and ergonomic scores from the ISONORM 9241/10 questionnaire.

the grab gesture is natural and logical, they had problems with the continuous 3D rotation. In the case that the rotation was restricted by their hand joints, they had to release the object, place the hand to a new position, grab again and rotate further. Then, it occurred that the object was not released correctly and a false-positive grab gesture was performed, which led to frustration. Relating to this effect, there were ambiguities when zooming: when changing the hand posture from the *open*-hand rest gesture to the closed *fist* gesture, at some point the finger positions are similar to the *grab* gesture, thus performing an unwanted rotation and leading to frustration and delays. The swipe gesture, however, is well known from handheld devices and therefore easy to use and natural for most users. For this reason, the swipe gesture is recommended to use for 3D object rotation if a coarse rotation is sufficient. The acceptance of a discrete rotation gesture could be increased if the angle steps were adjustable and not fixed. A surprising aspect is that some participants were not able to perform all gestures correctly due to restrictions in motor coordination. Especially the continuous 3D rotation was problematic, although it was inspired by grabbing and rotating a real object. Because some users were able to adapt to the restrictions of the grab gesture more quickly than others, the training times differed noticeably. Thus, the gesture for the continuous 3D rotation must be refined to be more robust, less tiring, and accessible for all users.

The UI in total was rated suitable for the task, self-descriptive, and uniformly designed. However, there is need for improvement in error tolerance, controllability, and the amount of help needed to learn the usage of the UI (see Figure 5.6). Hence, the false-positive rate of the gesture recognition must be reduced – possibly by integrating a lock/unlock voice command or gesture. An interactive tutorial could introduce the gestures and GUI to the users so they are able to discover the functions on their own. In addition, the prototypical hardware setup needs to be integrated

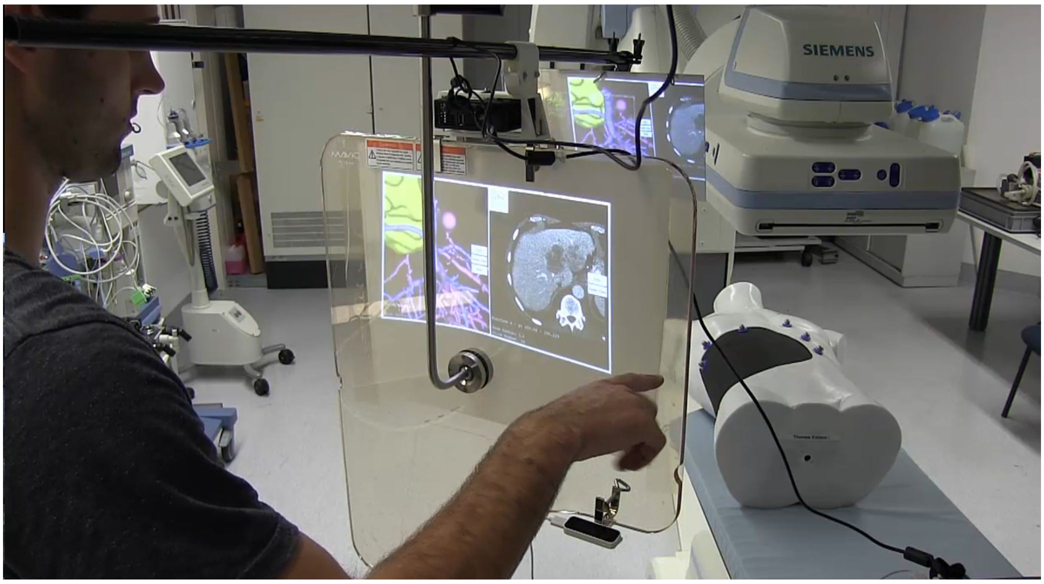


Figure 5.7: Projector and LMC mounted onto the radiation shield with a custom 3D-printed mount.

better. Therefore, a compact mounting of the projector onto the radiation shield is required.

If applied as suggested, the proposed prototype of a touchless gesture-controlled interactive radiation shield can lead to a more efficient workflow, better ergonomics for the radiologist, and safer interaction with medical images during CT-guided interventions.

5.1.7 Follow-Up

With the experience from the user study, some of the suggested improvements were made. First, a mount for attaching the projector onto the radiation shield and for the LMC was 3D printed. These are shown in Figure 5.7. This way, the display stays functional when moving the radiation shield.

Second, the gesture set was partly redesigned to resolve ambiguities. Therefore, the *swipe* and *grab* gesture were discarded and the *open* hand gesture was mapped to the 3D rotation. When extending the fingers, the current palm normal vector facing downwards is taken as a reference from which an angle difference is calculated when tilting the hand. The magnitude of the angle and direction of the hand tilt map onto the magnitude and direction of the rotation. This behaviour is metaphoric to rolling an object along an inclined plane. The new rest gesture is a relaxed hand with the fingers slightly bent, so that no ambiguities and thus not unintended gestures occur anymore. The new gesture set[1] was evaluated in comparison with interventional touch screen input in a user study [179]. Here, the performance of the rotation gesture is comparable to dragging on a touchscreen, although the participants had much more experience with touchscreens than with 3D gesture interaction. For the other gestures, slightly worse performance than touch screen interaction was measured. It became apparent that more complex interaction, such as 3D rotation, can be better applied with input with more degrees of freedom.

To fully leverage the potential of touchless gesture interaction and interact completely sterilely with software in the intervention room, the gesture control should be applied to all software involved in CT-guided interventions. It is even possible to extend the use case of this type of intervention to many other intervention scenarios, such as MRI-guided interventions, because the gestures are designed for abstract functions, such as selecting, panning, zooming, and rotating.

5.2 Touchless Gesture Control of an MRI Scanner

In the Chapters 1, 2 and 4, the advantages of MRI-guided interventions were explained and an AR navigation system for in-bore usage was presented that provides needle guidance support for percutaneous tumour ablations. However, during the user studies, it became obvious that radiologists must still rely on live MR images showing the correct relation between the applicator, the target, and risk structures. This is even more important when the liver and other organs move due to breathing.

Live images acquired with the MRI are usually displayed on a separate screen next to the patient table by mirroring the MRI control software. This software is intended to be operated by conventional hardware buttons, a trackball, or a computer mouse disregarding the clinical need for sterility [229] and worsening an already unergonomic situation for the interventional radiologist (see Figure 5.8). Additionally, from a usability point of view, the software is designed for diagnostic requirements and is thus overloaded with functions that add complexity to its control while not generating benefit during interventions. Thus, the radiologist delegates the interaction with the MRI to assistants, which suffers from the issues as described in the introduction of this chapter.

Therefore, an adapted version of the gesture control presented in the previous section allows the radiologist to operate the MRI consistently, easily, and directly during interventions from inside the scanner room. Touchless hand gestures are translated to functions of a custom-designed interventional software. Both the software and the gestures are designed to efficiently and unambiguously start,

[1] A video demonstration is available at `https://archive.org/details/InteractiveRadiationShieldNUI`.

Figure 5.8: Two radiologists in the scanner room (seen through a shielded window). One is performing the intervention inside the bore while monitoring the ablation needle pose on the MRI real-time images on the distant in-room screen. The other is assisting.

stop, and switch MRI sequences, adjust the image windowing, and interactively move the image planes in the three-pane-view to match the applicator's position and alignment. The sequences are preconfigured for the specific requirements of percutaneous liver tumour ablations but may be easily adapted in order to target other types of MRI-guided interventions.

After describing existing concepts for touchless control of the MRI, the workflow of MRI-guided percutaneous ablations in the liver is analysed with a focus on HCI. Then, the hardware, software, and interaction of the new touchless scanner control is presented. Furthermore, the results of the user study in a clinical setting comparing the touchless approach with the clinical state of the art, task delegation [13, 19], with regard to usability and subjective mental workload are shown and discussed.

5.2.1 Related Work on MRI Gesture Control

Touchless input methods have been researched for several years, motivated by sterility concerns with conventional devices such as mouse, keyboard and touchscreens [229, 251] (see Section 5.2.1).

These approaches cannot easily be applied to the iMRI. Because of the strong magnetic field, special MRI-conditional hardware must be used in order to not endanger patient or physician and to ensure that no additional noise is introduced to the imaging. Moreover, these are not tailored to the tasks that are performed during an MRI-guided intervention. Interactive, direct plane adjustment was realised by using manipulators [70], a wireless active tracking array based on single sideband amplitude modulation [252], in-bore optical tracking [71], or handheld devices with inertial sensors activated by foot pedals [253].

Güttler presented a proof of concept for the touchless control of an MRI scanner using hand gestures. An MR-conditional RGB camera recognises the number of extended fingers and the hand position. The system allows for moving the FOV, tilting the acquisition planes by 90° and starting the image acquisition [254]. However, this approach lacks a natural gesture set, a graphical user interface suitable for gesture interaction as well as a decent evaluation.

Rube et al. aimed to reduce delays during MRI-guided percutaneous interventions caused by the need to leave the scanner room for communication and image viewing [255]. They developed a web-based UI on a tablet PC which enables the radiologist to directly control the MRI. The functions include switching the sequences, changing the scan geometry and other related parameters. Nevertheless, this concept does not take into account sterility issues that arise with the use of touchscreens [251] or usability problems when the touchscreen is covered with a sterile drape.

5.2.2 Requirements and Setup

In order to better understand the needs and requirements for interaction with software during MRI-guided interventions a workflow analysis and an online survey were conducted. From the results, the functions of the prototypically implemented interventional MRI control software were derived. These functions are assigned gestures which were previously designed, evaluated (see Section 5.1, and improved (see Section 5.1.7). The details are described in the following.

Requirement Analysis

The requirements were analysed in two steps. First, issues in terms of HCI during MRI-guided interventions were identified by observing an MRI-guided tumour ablation of the liver and discussing the workflow with the clinical partners. The intervention followed the workflow described in Section 2.2.1. The observation revealed what was already described in the introduction: interventional interaction with the MRI and the acquired images was completely delegated to assistants. The trackball and buttons to operate the GUI on the in-room screen were not only out of reach most of the time (as shown in Figure 5.8) – they were not even functioning. As for the interaction, the following frequently used tasks were identified:

- displaying planning data,
- manipulating slice position and orientation,
- switch between plane orientation modes,
- control sequences and change sequence parameters,
- windowing.

Second, an online survey among international interventional radiologists was conducted to verify and quantify the findings. Twelve radiologists participated in the survey, two of them partly. The mean experience with needle interventions over all participants is 14.6 ± 8.4 years, the mean experience with MRI-guided interventions is 13.4 ± 7.9 years and the mean number of performed needle interventions is 603.2 ± 752.2. The questionnaire consisted of four questions. The first one relates to the state-of-the-art method to control the MRI during interventions (see Figure 5.9), while the last three questions assess the pre- and interventional use as well as the potential usefulness of a set of functions (see Figures 5.10 and 5.11).

Based on the survey results, the following functions were determined that are essential for supporting MRI-guided interventions and implemented in the software:

- Starting/stopping/changing sequences

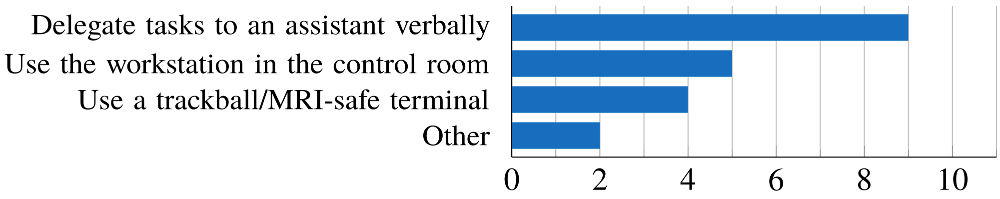


Figure 5.9: Survey results on "How do you currently control the MRI scanner?". Multiple selections were possible.

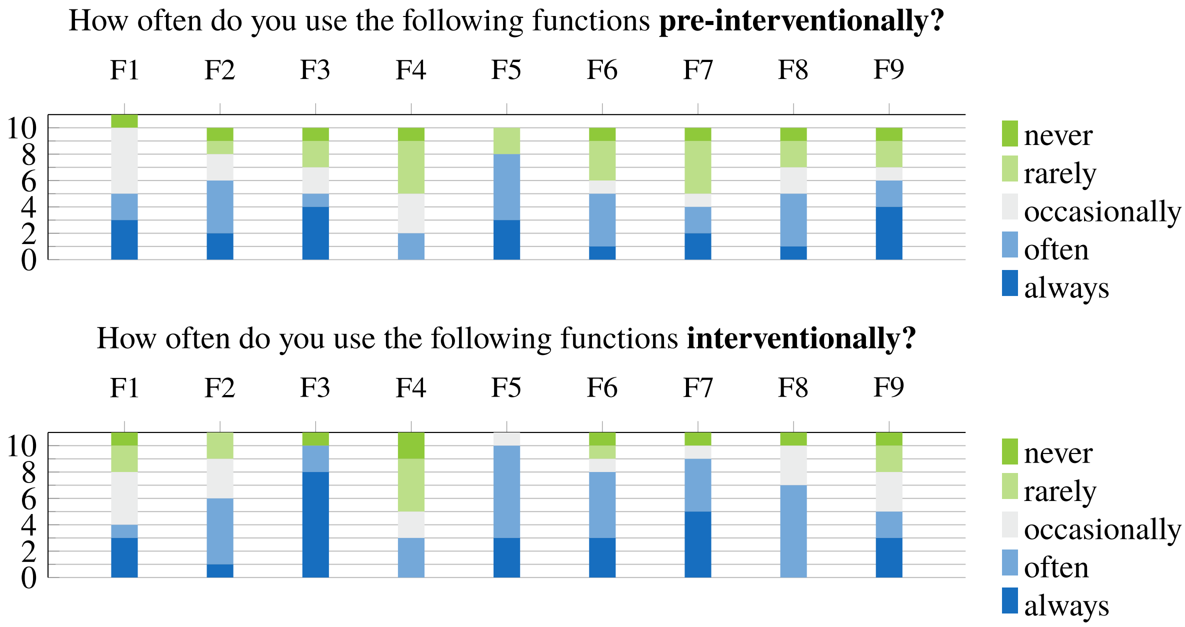


Figure 5.10: Survey results on usage of the following functions before and during interventions:
F1 - Windowing
F2 - Change sequence
F3 - Start/stop sequence
F4 - Change sequence parameters
F5 - Translate image plane
F6 - Rotate image plane
F7 - Switch between planes that are parallel or orthogonal to the needle
F8 - Switch between sagittal, coronal and axial plane
F9 - Show planning data sets (including tumour in 3D, entry point, planned trajectory, target)

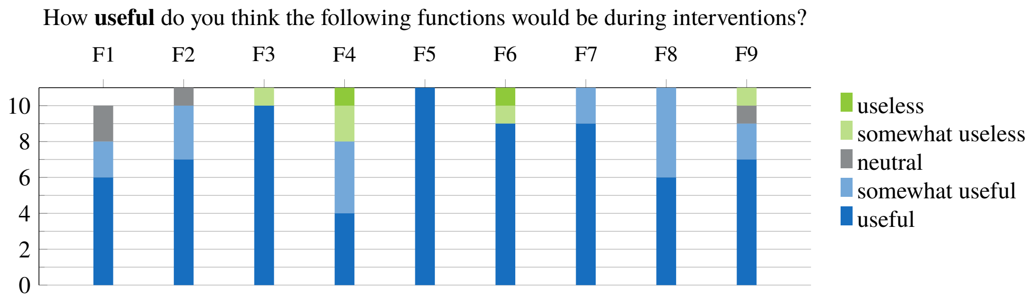


Figure 5.11: Survey results on interventional usefulness of certain functions. The functions are the same as in Figure 5.10.

- Windowing
- Moving (translating) image planes

User Interface

Based on the results of the requirement analysis, a prototypical interaction system was created that allows for direct control over the MRI from within the scanner room. It is operated with touchless hand gestures to maintain sterility. The system consists of two main parts. The first one is the software that provides the frequently used functions identified before on a GUI especially designed for gestural input. Second, an intermediate device was created to transmit the data from an off-the-shelf hand gesture sensor (LMC) to the client computer without interfering with the MRI.

The main purpose of the software is to display real-time images on the MRI-conditional in-room monitor and to provide a graphical user interface for scanner control using hand gestures. Respecting the de-facto standard image view radiologists are used to, the proposed GUI displays three image planes either orthogonally oriented along the planned applicator path or parallel to give contextual information on the surrounding structures while progressing towards the tumour (see Figure 5.13). Contrary to a linear, diagnostic workflow, interventional MRI sequences are required to provide rapid image acquisition rates. This leads to a smaller set of sequences to choose from. Thus, three preselected MRI sequences are provided via the prototypical software interface used, each in two parameter settings favouring speed or image resolution (*fast* and *high quality*).

The UI is controlled by different types of hand gestures. The pointing gesture described in Section 5.1.3 controls the cursor by absolutely mapping the space coordinates over the gesture sensor to the screen. By performing the click gesture (see Section 5.1.3) a selection is triggered (see Figure 5.12a-b). To be able to interactively move the image slice positions, the open hand gesture (see Section 5.1.3)

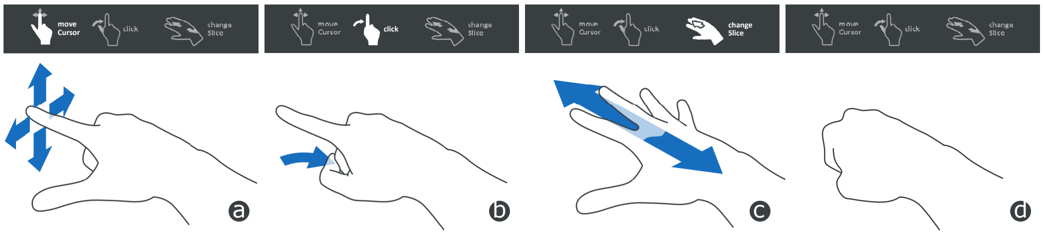


Figure 5.12: Hand gesture set for touchless MRI control: moving the cursor (a), triggering a selection (b), moving the image slice position (c), no interaction (d). Within the software user interface, an icon bar indicates the currently detected gesture. The corresponding state of these icons can be found above each gesture. For gesture (c), the icon highlights extra states for forward, backwards and no movement. ©2019 Springer Nature

was adapted to semantically connect to image plane translation (see Figure 5.12c). An open hand initiates the null-position for slice movement. Moving the open hand to the front or back translates the slice position at a rate proportional to the distance to the null-position. The translation is stopped upon changing the hand gesture or leaving the sensor area. An indicator bar on top of the screen provides visual feedback on the detected gesture (see Figure 5.13).

The commercially available Scanner Remote Control (SRC) programming interface (Siemens Healthcare GmbH, Erlangen, Germany) is used to directly control the MRI scanner. It allows to load, start, stop and change sequences, including adjustment of plane positions as well as changing all parameters available over the original GUI via an https-based REST interface.

The LMC is not MRI-compatible. Therefore, a custom solution was found by Pannicke et al. [256] to make it MRI-compatible. To this extent, the unshielded USB cable was exchanged with a fibre optic cable for data transmission and a battery with a charge controller for power supply, which lasts approximately 5 h. To avoid the induction of voltage spikes in the electronics, a surge filter (Würth Electronic, Niedernhall, Germany) which keeps the voltage below a safe threshold was added. These components were enclosed in an aluminium housing. The cable connecting the housing and the LMC was shielded. This setup ensures that the components remain functional during image acquisition and do not introduce a significant amount of additional noise to the MRI.

5.2.3 Evaluation

To evaluate the direct interaction concept, the touchless gesture input was compared to task *delegation*, which is clinical practice according to the online survey (see

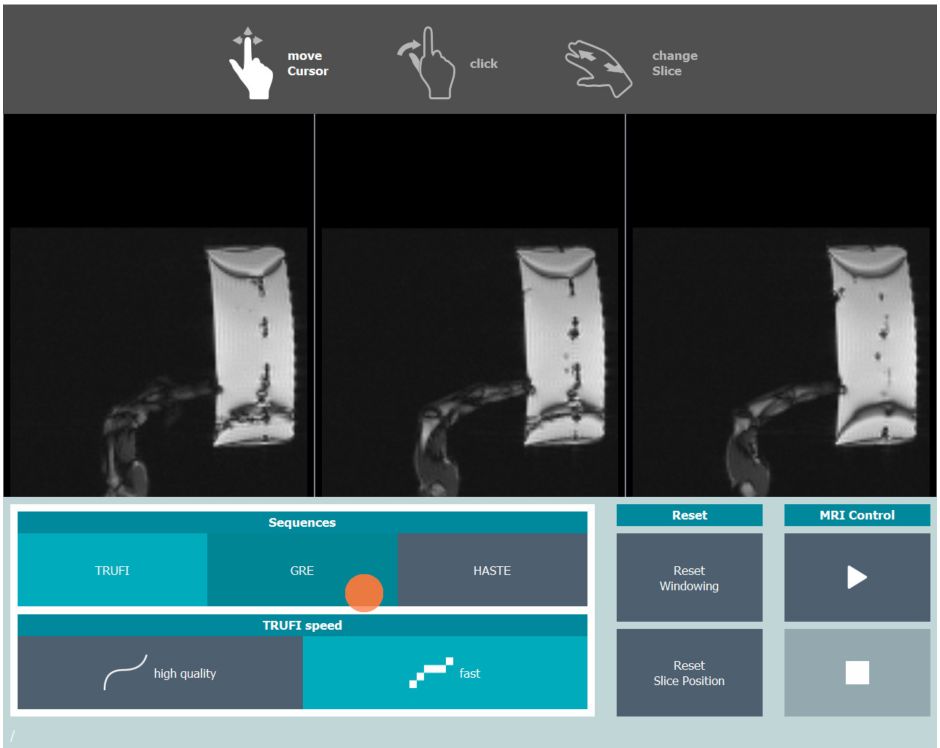


Figure 5.13: GUI for direct, touchless MRI control. A set of imaging sequences can be found at the bottom bar on the left, MRI control elements are on the right and visual feedback for hand gesture recognition is shown at the top. ©2019 Springer Nature

Figure 5.9). Only physician-computer-interaction inside the MRI scanner room was evaluated. Additional steps such as trajectory planning, which could be performed on a conventional workstation beforehand, were omitted.

Participants

Ten radiologists (six male, four female) participated in the study. Their age ranged from 27 to 50 (33.4 ± 7.0); average experience with MRI-guided interventions was 0.35 years (MIN=0,MAX=2,SD=0.75); experience with needle interventions ranged from 0 to 18 years (4.4 ± 6.1 years), mostly with CT interventions (3.8 ± 6.3). All of the participants were right-handed.

Apparatus

The user study was conducted using a 1.5 T MRI scanner (MAGNETOM Aera, Siemens Healthcare GmbH, Erlangen, Germany) controlled by a workstation running the Siemens Dot Cockpit Software. The setup was arranged for participants standing in front of the MRI next to the bore to provide a good view of the MRI-compatible workstation on the opposite site of the patient table and the control room window, similar to an interventional setup (see Figure 5.14). A phantom consisting of a plastic box filled with candle gel and three rubber o-rings serving as targets near the bottom was placed inside the bore. In front of the participant at the side of the patient table, the gesture sensor was placed and connected with a PC in the control room. The PC was connected to the MRI host via Ethernet cable and used to control the MRI via SRC and provide the proposed scanner control UI.

Depending on the input method, the manufacturer's workstation or the PC were connected to the in-room display next to the MRI scanner.

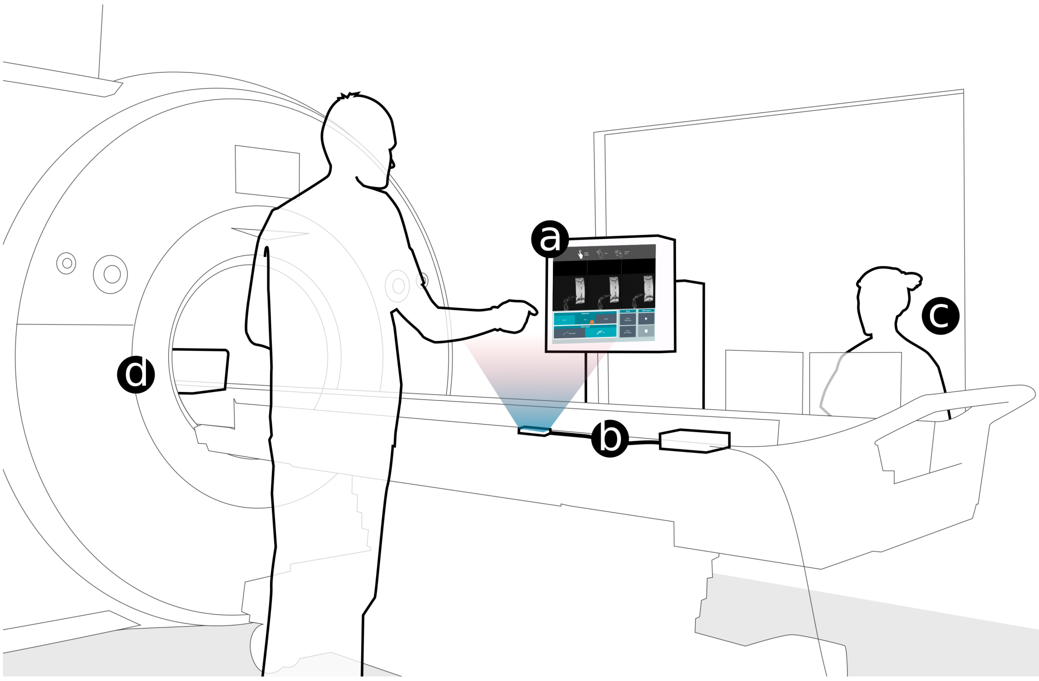


Figure 5.14: Study setup consisting of an MRI-compatible monitor (a), an MRI-compatible gesture controller (b), an assistant in the control room (c) and a candle gel phantom inside the MRI (d). ©2019 Springer Nature

Study Design

The user study followed a within-subject design. Two interaction modalities (touchless *gesture input* and task *delegation*) were used to perform the six tasks described in Section 5.2.3. The order of input methods was randomised.

Tasks

The set of tasks for the users to be performed aimed at emulating the workflow of an MRI-guided percutaneous tumour ablation. It consisted of the tasks "*Start* Sequence", "*Move* Slice Position", "*Window* to Given Width and Level" and "Insert *Needle* to Target". Because the *Start* task was expected to take only a few seconds, it was executed three times per modality and user to level out possible short delays during the task execution (either by the user or the MRI). The goal of the first three tasks was to change a running sequence to another, predefined sequence. All three of these *Start* tasks required the same number of actions to achieve the goal. Before the beginning of the first task, the study operator started a sequence to ensure equal conditions for all tasks. To change the sequence, the user needed to stop the current sequence, choose the desired one in the specified parameter variant (*fast* or *high quality*) and start this sequence. In the *Move* task, the sagittal slice group was to be translated along the left-right-axis over a distance of 50 mm to a defined slice position. The task *Window* required the user to adjust window width and level to a desired value each. Due to the difficulty of pinpointing the exact values, a tolerance of 100 for the width and 60 for the level was introduced. During the final task *Needle*, no interaction with the MRI was necessary. Only the needle was to be inserted into the phantom relying on the real-time MR images. The purpose of the needle insertion task was to provide a complete picture of the workflow, which helped less experienced participants give feedback on the differences and improvements between both input methods. All tasks were to be executed in the described order.

The tasks differed slightly in execution between the two modalities. While the users were able to perform all tasks directly with *gesture input*, an indirect path had to be taken when *delegating*. The user needed to attract the attention of the assistant in the scanner room and then communicate the commands via the intercom. Due to the loud noise, the running sequence had to be stopped by the assistant before verbal communication was possible. The assistant confirmed all commands via intercom and performed them in approximately the same speed for all users in the Siemens Dot Cockpit software. During the *Move* task, the only commands allowed were "left", "right", and "more". The assistant then moved the slice group accordingly. To *window*, the users had to make the assistant stop the running sequence and come out of the scanner room. Then, the assistant was told to adjust the window width "wider" or "narrower" and the window level "brighter" or "darker" until the target values were met. This best approximated the state-of-the-art workflow.

Measures

As a measure of performance, the task completion time (TCT) was taken. Task durations were logged automatically via the MRI software interface SRC except for the windowing task when using task delegation and the needle placement task, which were gathered manually. Subjective workload was assessed with the NASA Task Load Index questionnaire (NASA-TLX) without weighting of the subscales, also known as Raw TLX [218] and the System Usability Scale [257]. Subjective usability was measured using the System Usability Scale (SUS) [257]. A post-test interview was conducted at the end.

Procedure

The study took place in the MRI Suite of the Hannover Medical School outside business hours. Two study supervisors were present during the entire study. One of the supervisors instructed the participants in the scanner room. The other remained in the control room, prepared the software, and played the role of an experienced assistant for the modality of *task* delegation. For each user, a phantom was prepared by the operators, with the needle already placed at the entry point. The supervisor then described the entire procedure and the required workflow steps. With touchless *gesture input*, all tasks were performed next to the MRI scanner (see Figure 5.14), whereas *delegation* was tested as described in Section 5.2.3. The same protocol was followed for both methods: At the MRI scanner, the first of the two input modalities was explained by the instructor. Every kind of interaction was demonstrated by the instructor and questions were answered. The participants were given unlimited time to familiarise themselves with the *gesture input*. Then, the six tasks were explained by the instructor. A short optical sign indicated the beginning and the end of each task. After completing all tasks with an input method, the participants completed the NasaTLX and the SUS questionnaire. The same procedure was then followed for the second input method. Finally, the expert interview was conducted.

5.2.4 Results

All participants successfully completed all tasks. Due to a software error, the log files of the first user were not recorded. Therefore, the quantitative measurements (task durations) are based on 9 users only, but the qualitative measurements (NASA-TLX, SUS, interview) contain the data of all 10 participants.

The users needed on average 360 ± 117 s until they felt confident to perform the various tasks. The task completion times of the tasks are shown in Figure 5.15. It is illustrated that the completion time of the *Start* task is the same for both modalities. The same applies to the *Needle* task, which was the same for both modalities,

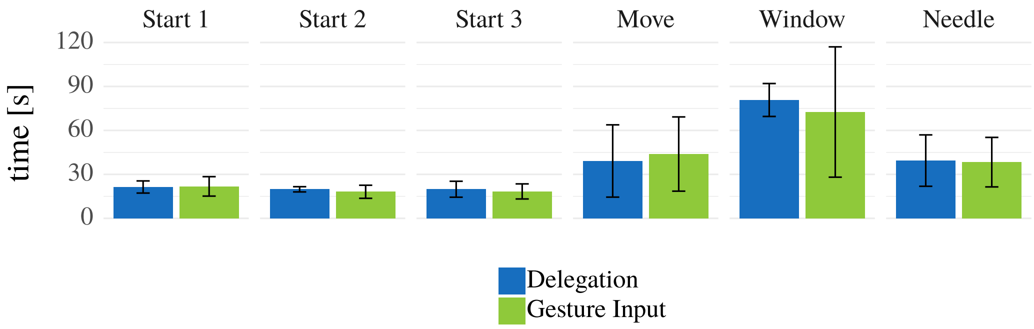


Figure 5.15: Task duration of the six tasks with *Start* 1/2/3 being the "Start sequence X" tasks, *Move* the slice positioning task, *Window* the windowing task and *Needle* the needle guidance task.

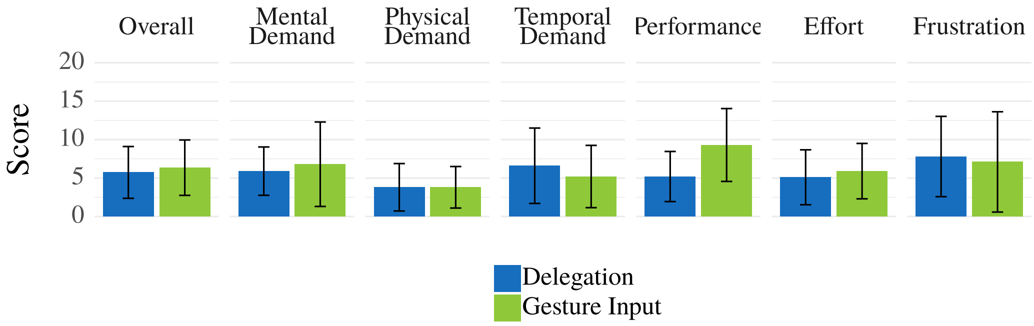


Figure 5.16: Subjective workload score from the NASA-TLX questionnaire.

because no software interaction was required. Differences in task duration were observed in the *Move* and *Window* tasks. The slice position shift with *gesture input* was slightly slower (43.9 ± 25.3 s) than *delegation* (39.1 ± 24.7 s) of the task, but *Window* was faster with *gesture input* (72.5 ± 44.5 s vs. 72.5 ± 44.5 s), albeit with greater deviation.

The NasaTLX scores are shown in Figure 5.16. The total score and the perception of mental and physical demand as well as effort and frustration are comparable with both modalities. However, there are differences in the valuation of temporal demand and performance. The temporal demand is perceived higher with *delegation* (6.6 ± 4.9) than with *gesture input* (5.2 ± 4.0). The performance is subjectively rated worse with *gesture input* (9.3 ± 4.7 vs. 5.2 ± 3.6).

Comparing the SUS scores, *gesture input* scores higher in usability (74.8 ± 14.9) than task *delegation* (58.3 ± 25.4).

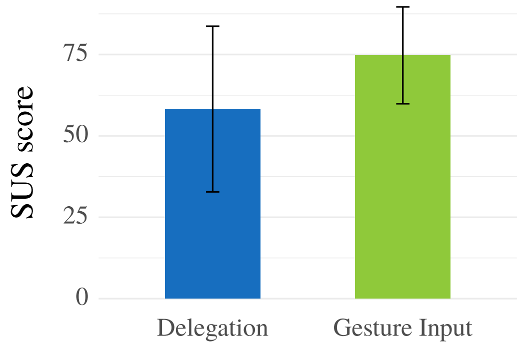


Figure 5.17: System usability derived from SUS questionnaires.

Nine of the ten participants mentioned that it is a great advantage to have direct control because it is more predictable than an assistant, and because it enables decision-making to be more free and at a higher level of control. One user pointed out that interventional radiologists are used to directly control the table movement of a CT angiography suite without the help of an assistant, making the *gesture input* approach more familiar. Another admitted to use a gesture-controlled system rather than performing extensive hand gestures against the assistant in front of a conscious patient to maintain a professional atmosphere. Six participants assumed that the *gesture input* saved time because it was not necessary to leave the scanner room or interrupt the task, perform additional verbalisations, or wait for the assistant's response.

Four users described the disadvantages that interaction via a proxy user could have, such as assistants without much experience, dependence on someone else who relies on error-prone verbal communication, or a lack of necessary precision. Three participants had difficulties using different 3D planes for directed gestures: One user thought it would be easier to use the same plane for windowing and plane adjustment, while two others suggested choosing gestures more similar to mouse movements on a desktop. Although there was visual feedback for the gesture currently detected, three participants reported a missing indicator for the tracking volume of the gesture sensor. One user reported fatigue in the raised arm during gesture interaction. Three participants expressed concerns about integration into the current workflow and working environment. It was suggested to place the gesture sensor in front of the physician and not next to the patient. A radiologist explained that assistants outside the operating room are not kept informed about progress, making the procedure difficult to follow for assistants, observers, or students. This makes it difficult to prepare in time for supportive tasks. Gesture interaction was considered a possible distractor when needle placement was performed by one radiologist.

5.2.5 Discussion

In this section, the setup and evaluation of a touchless gesture control for the iMRI has been presented which allows radiologists to directly interact with the scanner and the image visualisation. To best integrate the control into the existing workflow, requirements from the literature, the observation of interventions, information from clinical partners, and an online expert survey were collected. Based on the results, the functions of the system were implemented and evaluated against the *delegation* of interaction in a user study.

Hand *gesture input* produced comparable results as the task *delegation* when it comes to the completion time of the task and the subjective workload (see Figure 5.15 and 5.16). However, the times measured during the *Move* task are slightly higher for the *gesture input* than for the *delegation*. One of the reasons for this is that many users first moved the slices in the wrong direction. Most of them were unaware that the start value was left, but the target value was right in the patient coordinate system, leading to a misinterpretation of the target position.

The subjective performance was evaluated for the use of touchless *gesture input* lower than for the task *delegation*. However, although there is greater variability for the *gesture input*, the average task completion time does not reflect such a difference.

In terms of subjective usability, the SUS score shows an average rating of 74.8 for hand *gesture* interaction and 58.2 for verbal task *delegation* (see Figure 5.17). These results can be interpreted in terms of acceptance and adjective rating (worst imaginable, worst, ok, good, excellent, best imaginable) [258]. The rating for hand *gesture* interaction corresponds to an acceptable performance and lies on the adjective scale between "good" and "excellent". The evaluation of the verbal task *delegation*, on the other hand, means low marginal acceptance and lies in the range between "ok" and "good". These results indicate a high level of acceptance and give indications of further improvements in order to exceed the verbal task *delegation* in terms of usability and performance in the future. By providing a suitable physician-computer interface, this work has the potential to improve the procedures guided by MRI.

Although the study was carefully designed, various factors come into play in clinical practice that are difficult to consider in terms of study design. The task *delegation* strongly depends on the assistant's experience, knowledge of the intervention and the workflow. There is no standardised vocabulary, i.e. it depends on the preference of the physician. In the worst case, external events require the assistant's attention, making interaction temporarily impossible. Moreover, comparison with the literature is difficult because similar study designs show different results that could be caused by the way the task delegation is performed [167, 259]. The working range of gesture control was defined in the study, but should be personalised for clinical use. In addition, the space available in the bore and on the table varies and requires a more flexible approach to sensor placement. The unfamiliar mapping of mouse input and image slice positioning at different levels in 3D space was uncovered in the post-test interviews. Despite the visual feedback on gesture recognition, unintended input occurred when the user had not paid close attention to when the "virtual stop" was released. This could be overcome in the future by using some kind of clutching mechanism [260].

5.3 Conclusion

In this chapter, a touchless, natural, and metaphoric hand and finger gesture set was presented that is intended be used for interaction in sterile interventional medical scenarios, especially with medical images. In order to integrate well into the intervention room environment in spite of shortness on unoccupied space, the radiation shield of a C-arm CT suite protecting the radiologist from X-rays was equipped with a projector and an LMC for hand and finger gesture recognition. Unlike most previous approaches that merely use hand postures or arm gestures, the proposed gesture set consists of deictic and iconic gestures. These are used to point and click, scroll through image slices, 3D rotate anatomical surface models, and scale and pan the image data. The UI was designed based on the guidelines of NUIs to maximise memorability, self-describability, and usefulness. The user study with 12 participants confirmed the suitability of the system for the intended tasks in general as well as a good gesture design for most gestures. However, weaknesses and ambiguities with the 3D rotation gestures were revealed that were addressed in a study follow-up. The rotation gestures were replaced with the simpler *open* hand gesture, which does not interfere with the *grab* gesture used for zooming and is not limited by the motor coordination of the users. The improved gesture set was rated better in a follow-up comparison study, but still performed worse than the compared touch screen interaction [179]. One could now draw conclusions about the general suitability of these gestures based on such results, but it has to be kept in mind that interaction with a touchscreen covered with a plastic drape when wearing rubber gloves is still cumbersome. Touchless gesture control is superior in terms of sterility, compared especially to pseudo-sterile practices such as covering a joystick with surgical gown [168] so that it is not touched directly, but also to plastic drape wetted with the patients blood and repeatedly touched. A possible better patient outcome may outweigh the missing haptic feedback and performance issues of touchless gestures. Furthermore, most users have a lot more experience with touchscreen interaction than with touchless gestures. Touchless gestures provide more DOF, which is why more training may be needed to adapt to the modality.

Therefore, similar studies with better trained users should be performed in the future to be able draw a sound comparison between the input modalities.

In the second scenario, a part of the gesture set was adapted and integrated into an interventional UI used to control the MRI directly while performing an MRI-guided intervention. The software functions were defined after clinical needs and compared with interaction task delegation as the state of the art. Here, the gesture interaction in combination with the specialised interventional UI performed similar to current clinical practice in terms of task completion times and subjective workload, and even showed a slightly better usability. These results show that a carefully selected gesture set and functions adapted to the application play an important role when introducing a new, unconventional interaction modality. Before integrating a direct interventional interaction with the MRI for the radiologist into the clinical practice, it should be investigated how big the impact on the assistants is, who are less directly involved in the intervention and may be free for other supportive tasks.

To further reduce the recognition of unintended gestures, a multimodal approach may be followed by introducing voice commands to activate or lock functions and gestures. The interaction in total may be further minimised by automating different functions, such as the automatic plane alignment to the needle based on instrument tracking data as is done by Kägebein et al. [71].

The chapter is based on research that has been published in:

A. Mewes, P. Saalfeld, O. Riabikin, M. Skalej, C. Hansen. “A gesture-controlled projection display for CT-guided interventions”. In: *International journal of computer assisted radiology and surgery* 11.1 (2016), pp. 157–164. [178]

A. Mewes, B. Hensen, F. Wacker, C. Hansen. “Touchless interaction with software in interventional radiology and surgery: a systematic literature review”. In: *International journal of computer-assisted radiology and surgery* 12.2 (2017), pp. 291–305. [146]

B. Hatscher, A. Mewes, E. Pannicke, U. Kägebein, F. Wacker, C. Hansen, B. Hensen. “Touchless scanner control to support MRI-guided interventions”. In: *International Journal of Computer-Assisted Radiology and Surgery* (Sept. 2019). [261]

The contents of Section 5.2 and the third publication [261] are a joint work between Mr. Benjamin Hatscher, the first author, and the author of this thesis, whereby both authors contributed equally to the review of related works, the workflow observation and communication with the clinical partners, the conception, planning, performance, and evaluation of the user study as well as to the discussion. Mr. Hatscher conducted the online survey regarding clinical interaction practices, created the GUI and the connection with the MRI, and integrated the gestures, which were in turn conceptualised, tested, and initially implemented by the author of this thesis. The authors of the publication confirmed by mutual agreement that the work may be used by all of them within the main focus of each persons work, which, in this case, is the translation of the previously developed natural gesture set to the scenario of MRI-guided interventions to improve the workflow.

6 Conclusion

The MRI is a remarkable device, which enables physicians to gain insights into the human body without the need for radical, invasive surgery while providing very good image quality that is often superior to other imaging modalities, such as CT or US. Because of its high soft tissue contrast it is not only ideal to diagnose diseases in organs. Due to fast imaging capabilities in arbitrary image plane orientations the MRI has also become an attractive interventional imaging modality to support percutaneous interventions.

One disease regularly treated in the iMRI is liver cancer, either in the form of primary liver tumours or liver metastases. When there are only small and few nodules present [5] or even for bigger tumours [35], they can be treated with thermal ablation. In this case, a needle-shaped applicator is guided under real-time imaging to the tumour, where then heat is applied until – ideally – all tumour cells have died from coagulation necrosis and irreversible damage. Unfortunately, the process of finding the entry point on the skin and of the actual needle guidance is not supported properly. Few navigation solutions exist and their information is always displayed on a screen next to the MRI. In addition, the MRI control software is designed for diagnostic purposes with a lot of functions not needed during interventions thus adding complexity to the procedure and increasing the mental demand of the radiologist. The MRI is usually operated by an assistant, because the in-room controls are often out of reach and cannot be used sterilely, adding potential for confusion and frustration and finally delaying the intervention.

In this dissertation, different technologies have been presented to improve the current workflow of MRI-guided interventions, including the first projector-based AR navigation system to augment the operating field inside the MRI bore with needle guidance information, and a natural, direct touchless gesture control of the MRI during interventions.

In order to support the radiologist during percutaneous MRI-guided interventions, a projector was set up outside the MRI intervention room. Its light is guided through three mirrors onto the operating field inside the MRI bore. It is calibrated with a ceiling mounted camera and registered with the MRI as well as with a needle tracking system based on MP markers. The projector-camera system is used to 3D scan the projection surface as well as projecting navigation cues onto it. Therefore, two navigation concepts were developed: the first provides two-dimensional explicit visual aids on the entry point, target, and needle pose and depth, the second consists of a three-dimensional virtually elongated needle representation and correspondingly a 3D entry point and target.

The user study confirmed a high accuracy with a target error comparable to the state-of-the-art freehand method reported in the literature. This error can be taken as a measure for not only the users' performance during the study but also for all accumulated errors from the calibration. The reprojection errors of the camera, the projector, and the stereo calibration were less than 1 px each indicating an accurate calibration. High meCUE and low NasaTLX scores confirm the ease of use of the navigation concepts with clear feedback even without a lot of training or experience with either AR or needle interventions.

However, these good results have so far only been achieved with the help of a few assumptions and simplifications that have to be ruled out in order to translate the AR system to clinical application. The AR projection does not consider any movement of the projection surface yet, so that a breathing patient would cause the insertion point not to be displayed at the correct position. This needs to be considered in the future, e.g. by applying and tracking markers or a marker pattern on the body by their possible integration into the incision foil in order to update the 3D-scanned projection surface similar to Narita et al. [225]. In addition, no organ deformations are taken into account yet. This can lead to deviations of the displayed navigation information from the real situation, which increases the risk of injury to risk structures. In addition, the target cannot be reached reliably in this way. Therefore, fast registration methods should be applied on the live MRI images in the future to update the navigation data in real-time. Otherwise, the radiologists must still use the separate in-room monitor to ascertain that the needle is on the correct path. Possible solutions are the generation of a predictive breathing motion model based on live imaging and a pre-interventional high resolution scan proposed by Xu et al. [224], or generating an elastic 3D model from pre-interventional data and update the target position with interventional live-imaging introduced by Westwood [223]. The needle tracking should further be improved to cover a larger volume so that all occurring needle poses during an intervention may be tracked. It should be further investigated, whether and how a precise viewpoint tracking of the user can increase the depth perception of the rendered scene with kinetic depth cues if 3D visualisations such as risk structures or the tumour are shown.

The projector-based AR system is a major contribution to the iMRI, because it can replace the inaccurate finger tipping method to find the entry point and provide an orientation help even when the needle artefact cannot yet be seen in the live images on the first centimetres of the insertion. This may lead to less correction movements, less damage to healthy tissue, and faster progression of the intervention. The design of the AR setup is general enough that other types of interventions with a similar workflow can be explored.

To enable the radiologist to interact directly and, more importantly, sterilely with interventional software, such as medical image viewers or imaging device control software, a natural metaphoric hand and finger gesture set has been developed and evaluated in two scenarios. In the first scenario, the radiation shield of an angiographic CT-suite has been made interactive by projecting a software onto it that allows the radiologist through an NUI to scroll through image slices, zoom and pan the images, make selections with a cursor and rotate 3D surface models of anatomical structures. The gestures are recognised with an LMC placed below the radiation shield.

In a user study the gesture set was valued suitable for the interventional interaction, well-usable, self-describable, and easy to learn. Only the 3D rotation showed ambiguities and caused unintended input due to its similarity to the zoom gesture. In a follow-up, the gesture set was optimised and the rotation gesture was replaced in order to eliminate the ambiguities. A follow-up study [179] confirmed the improvements, but the touchless gesture control still performed worse in terms of task completion times than the compared touch screen interaction with medical images. However, touchless interaction has the advantage of sterile interaction, while touchscreens are covered by a surgical drape and touched repeatedly, which leads to an increased infection risk for the patient.

In future work, the missing haptic feedback must be levelled out by better visual or acoustic feedback that provides clear information on the current state of the input and the possible functions to be used. Unintended gestures should be avoided by introducing a trigger gesture [164] or, when following a multimodal approach, voice commands as function selectors [189]. In such a multimodal approach, the hand and finger gesture set may as well be combined with foot interaction or eye tracking to provide hands-free interaction when needed and seamlessly choose between different viewports to interact with [152].

The presented natural and metaphoric gesture set and its careful integration into the interventional scenario based on the design guidelines of NUIs is a promising step towards sterile and direct interaction and an improvement of the workflow of C-arm CT-guided interventions that eliminates the need to delegate interaction with medical images and that may well be translated to other types of interventions.

This has been done in the second scenario, in which an adapted version of the previously proposed gesture set was used to control a custom-designed NUI to directly operate the MRI and interact with acquired images during MRI-guided interventions. The software provides an NUI to select, start and stop predefined sequences in different parameter sets as well as to display and manipulate parallelly displayed live MRI images. It is shown on the in-room monitor next to the patient table of the MRI. The functions are based on clinical needs and have been determined beforehand.

In a user study with experienced radiologists, the gesture control performed comparable to task delegation to an assistant in terms of task completion times and a higher usability score despite the relatively short training time each user needed. The good rating of the gesture control is again based on the good integration into the intervention environment, the careful selection of the features, and the naturalness of the gestures in the sense of an NUI.

The results clearly show that there is a need for a better workflow for interventional operation of the MRI and interaction with medical images. The presented gesture interaction concept contributes to this issue by giving the radiologist the control over the tools needed to achieve the objective of the intervention without being mentally demanded too much, getting frustrated due to misunderstandings with the assistants, and in an ergonomic way.

In summary, the concepts presented in this dissertation contribute to a better workflow for MRI-guided percutaneous interventions and thus might support a more widespread adoption of this type of therapy. Beyond that, the projector-based AR system and the touchless gesture control of interventional software are universal in design, so that they may also be adapted to suit other areas of application, be it other types of interventions or even non-medical applications.

Bibliography

[1] P. Kucerova and M. Cervinkova. “Spontaneous regression of tumour and the role of microbial infection–possibilities for cancer treatment”. In: *Anti-cancer drugs* 27.4 (2016), pp. 269–277.

[2] J. F. Buell, M. T. Thomas, S. Rudich, M. Marvin, R. Nagubandi, K. V. Ravindra, G. Brock, and K. M. McMasters. “Experience with more than 500 minimally invasive hepatic procedures”. In: *Annals of surgery* 248.3 (2008), pp. 475–486.

[3] P. L. Pereira. “Actual role of radiofrequency ablation of liver metastases”. In: *European radiology* 17.8 (2007), pp. 2062–2070.

[4] K. Hasegawa, T. Aoki, T. Ishizawa, J. Kaneko, Y. Sakamoto, Y. Sugawara, and N. Kokudo. “Comparison of the Therapeutic Outcomes Between Surgical Resection and Percutaneous Ablation for Small Hepatocellular Carcinoma”. In: *Annals of Surgical Oncology* 21.3 (2014), pp. 348–355.

[5] P. R. Galle, A. Forner, J. M. Llovet, V. Mazzaferro, F. Piscaglia, J.-L. Raoul, P. Schirmacher, and V. Vilgrain. “EASL clinical practice guidelines: management of hepatocellular carcinoma”. In: *Journal of hepatology* (2018).

[6] P. Schullian, G. Widmann, T. B. Lang, M. Knoflach, and R. Bale. “Accuracy and diagnostic yield of CT-guided stereotactic liver biopsy of primary and secondary liver tumors”. In: *Comput. Aided Surg.* 16.4 (2011), pp. 181–187.

[7] G. Mauri, L. Cova, S. De Beni, T. Ierace, T. Tondolo, A. Cerri, S. N. Goldberg, and L. Solbiati. “Real-time US-CT/MRI image fusion for guidance of thermal ablation of liver tumors undetectable with US: results in 295 cases”. In: *Cardiovascular and interventional radiology* 38.1 (2015), pp. 143–151.

[8] L. Solbiati, T. Ierace, S. N. Goldberg, S. Sironi, T. Livraghi, R. Fiocca, et al. “Percutaneous US-guided radio-frequency tissue ablation of liver metastases: treatment and follow-up in 16 patients.” In: *Radiology* 202.1 (1997), pp. 195–203.

[9] M. D'Onofrio, R. De Robertis, E. Barbi, E. Martone, E. Manfrin, S. Gobbo, et al. "Ultrasound-guided percutaneous fine-needle aspiration of solid pancreatic neoplasms: 10-year experience with more than 2,000 cases and a review of the literature". In: *European radiology* 26.6 (2016), pp. 1801–1807.

[10] G. Belfiore, G. Moggio, E. Tedeschi, M. Greco, R. Cioffi, F. Cincotti, and R. Rossi. "CT-guided radiofrequency ablation: a potential complementary therapy for patients with unresectable primary lung cancer—a preliminary report of 33 patients". In: *American Journal of Roentgenology* 183.4 (2004), pp. 1003–1011.

[11] R. D. Suh, A. B. Wallace, R. E. Sheehan, S. B. Heinze, and J. G. Goldin. "Unresectable pulmonary malignancies: CT-guided percutaneous radiofrequency ablation—preliminary results". In: *Radiology* 229.3 (2003), pp. 821–829.

[12] K. Yasui, S. Kanazawa, Y. Sano, T. Fujiwara, S. Kagawa, H. Mimura, et al. "Thoracic tumors treated with CT-guided radiofrequency ablation: initial experience". In: *Radiology* 231.3 (2004), pp. 850–857.

[13] E. Rothgang, W. D. Gilson, F. Wacker, J. Hornegger, C. H. Lorenz, and C. R. Weiss. "Rapid freehand MR-guided percutaneous needle interventions: An image-based approach to improve workflow and feasibility". In: *Journal of Magnetic Resonance Imaging* 37.5 (2013), pp. 1202–1212.

[14] H. Rempp, H. Loh, R. Hoffmann, E. Rothgang, L. Pan, C. D. Claussen, and S. Clasen. "Liver lesion conspicuity during real-time MR-guided radiofrequency applicator placement using spoiled gradient echo and balanced steady-state free precession imaging". In: *Journal of Magnetic Resonance Imaging* 40.2 (2014), pp. 432–439.

[15] C. R. Weiss, S. G. Nour, and J. S. Lewin. "MR-guided biopsy: A review of current techniques and applications". In: *Journal of Magnetic Resonance Imaging* 27.2 (2008), pp. 311–325.

[16] M. A. Clifford, F. Banovac, E. Levy, and K. Cleary. "Assessment of hepatic motion secondary to respiration for computer assisted interventions". In: *Computer Aided Surgery: Official Journal of the International Society for Computer Aided Surgery (ISCAS)* 7.5 (2002), pp. 291–299.

[17] R. Hoffmann, C. Thomas, H. Rempp, D. Schmidt, P. L. Pereira, C. D. Claussen, and S. Clasen. "Performing MR-guided biopsies in clinical routine: factors that influence accuracy and procedure time". In: *European radiology* 22.3 (2012), pp. 663–671.

[18] A. Boss, H. Rempp, P. Martirosian, S. Clasen, C. Schraml, A. Stenzl, et al. “Wide-bore 1.5 Tesla MR imagers for guidance and monitoring of radiofrequency ablation of renal cell carcinoma: initial experience on feasibility”. In: *European radiology* 18.7 (2008), pp. 1449–1455.

[19] F. Fischbach, J. Bunke, M. Thormann, G. Gaffke, K. Jungnickel, J. Smink, and J. Ricke. “MR-guided freehand biopsy of liver lesions with fast continuous imaging using a 1.0-T open MRI scanner: experience in 50 patients”. In: *Cardiovascular and interventional radiology* 34.1 (2011), pp. 188–192.

[20] M. Bock and F. K. Wacker. “MR-guided intravascular interventions: Techniques and applications”. In: *Journal of Magnetic Resonance Imaging* 27.2 (2008), pp. 326–338.

[21] H. Busse, N. Garnov, G. Thörmer, D. Zajonz, W. Gründer, T. Kahn, and M. Moche. “Flexible add-on solution for MR image-guided interventions in a closed-bore scanner environment”. In: *Magnetic resonance in medicine* 64.3 (2010), pp. 922–928.

[22] S. G. Hushek, A. J. Martin, M. Steckner, E. Bosak, J. Debbins, and W. Kucharzyk. “MR systems for MRI-guided interventions”. In: *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 27.2 (2008), pp. 253–266.

[23] G. S. Fischer, A. Deguet, C. Csoma, R. H. Taylor, L. Fayad, J. A. Carrino, S. J. Zinreich, and G. Fichtinger. “MRI image overlay: application to arthrography needle insertion”. In: *Computer Aided Surgery* 12.1 (2007), pp. 2–14.

[24] M. Moche, D. Zajonz, T. Kahn, and H. Busse. “MRI-guided procedures in various regions of the body using a robotic assistance system in a closed-bore scanner: Preliminary clinical experience and limitations”. In: *Journal of Magnetic Resonance Imaging* 31.4 (2010), pp. 964–974.

[25] J. Weiss, R. Hoffmann, H. Rempp, D.-E. Keβler, P. L. Pereira, K. Nikolaou, and S. Clasen. “Feasibility, efficacy, and safety of percutaneous MR-guided ablation of small ($\pm 12mm$) hepatic malignancies”. In: *Journal of Magnetic Resonance Imaging* (2018).

[26] D. Manzey, S. Rottger, J. E. Bahner-Heyne, D. Schulze-Kissing, A. Dietz, J. Meixensberger, and G. Strauss. “Image-guided navigation: the surgeon’s perspective on performance consequences and human factors issues”. In: *Int J Med Robot* 5.3 (Sept. 2009), pp. 297–308.

[27] T. Akinyemiju, S. Abera, M. Ahmed, N. Alam, M. A. Alemayohu, C. Allen, et al. “The burden of primary liver cancer and underlying etiologies from 1990 to 2015 at the global, regional, and national level: results from the Global Burden of Disease Study 2015”. In: *JAMA oncology* 3.12 (2017), pp. 1683–1691.

[28] L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent, and A. Jemal. “Global cancer statistics, 2012”. In: *CA: a cancer journal for clinicians* 65.2 (2015), pp. 87–108.

[29] H. B. El-Serag. “Hepatocellular Carcinoma”. In: *New England Journal of Medicine* 365.12 (2011), pp. 1118–1127.

[30] A. Forner, M. Reig, and J. Bruix. “Hepatocellular carcinoma”. In: *The Lancet* 391.10127 (2018), pp. 1301–1314.

[31] M. M. Hassan, L.-Y. Hwang, C. J. Hatten, M. Swaim, D. Li, J. L. Abbruzzese, P. Beasley, and Y. Z. Patt. “Risk factors for hepatocellular carcinoma: synergism of alcohol with viral hepatitis and diabetes mellitus”. In: *Hepatology* 36.5 (2002), pp. 1206–1213.

[32] J. Ferlay, H.-R. Shin, F. Bray, D. Forman, C. Mathers, and D. M. Parkin. “Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008”. In: *International journal of cancer* 127.12 (2010), pp. 2893–2917.

[33] A. Gillams, N. Goldberg, M. Ahmed, R. Bale, D. Breen, M. Callstrom, et al. “Thermal ablation of colorectal liver metastases: a position paper by an international panel of ablation experts, the interventional oncology sans frontières meeting 2013”. In: *European Radiology* 25.12 (Dec. 2015), pp. 3438–3454.

[34] N. P. Malek, S. Schmidt, P. Huber, M. P. Manns, and T. F. Greten. “The diagnosis and treatment of hepatocellular carcinoma”. In: *Deutsches Ärzteblatt International* 111.7 (2014), p. 101.

[35] R. Bale, G. Widmann, and D. R. Stoffner. “Stereotaxy: breaking the limits of current radiofrequency ablation techniques”. In: *European journal of radiology* 75.1 (2010), pp. 32–36.

[36] K. F. Chu and D. E. Dupuy. “Thermal ablation of tumours: biological mechanisms and advances in therapy”. In: *Nature Reviews Cancer* 14.3 (2014), p. 199.

[37] S. P. Haen, P. L. Pereira, H. R. Salih, H.-G. Rammensee, and C. Gouttefangeas. “More than just tumor destruction: immunomodulation by thermal ablation of cancer”. In: *Clinical and Developmental Immunology* 2011 (2011).

[38] R. V. Davalos, L. Mir, and B. Rubinsky. “Tissue ablation with irreversible electroporation”. In: *Annals of biomedical engineering* 33.2 (2005), p. 223.

[39] M. Nikfarjam, V. Muralidharan, and C. Christophi. “Mechanisms of focal heat destruction of liver tumors”. In: *Journal of Surgical Research* 127.2 (2005), pp. 208–223.

[40] A. S. Wright, L. A. Sampson, T. F. Warner, D. M. Mahvi, and F. T. Lee Jr. “Radiofrequency versus microwave ablation in a hepatic porcine model”. In: *Radiology* 236.1 (2005), pp. 132–139.

[41] B. W. Dong, J. Zhang, P. Liang, X. Yu, L. Su, D. Yu, X. Ji, and G. Yu. “Sequential pathological and immunologic analysis of percutaneous microwave coagulation therapy of hepatocellular carcinoma”. In: *International journal of hyperthermia* 19.2 (2003), pp. 119–133.

[42] M. Ahmed, C. L. Brace, F. T. Lee Jr, and S. N. Goldberg. “Principles of and advances in percutaneous ablation”. In: *Radiology* 258.2 (2011), pp. 351–369.

[43] T. Mala. “Cryoablation of liver tumours–a review of mechanisms, techniques and clinical outcome”. In: *Minimally Invasive Therapy & Allied Technologies* 15.1 (2006), pp. 9–17.

[44] M. C. Jansen, R. van Hillegersberg, I. G. Schoots, M. Levi, J. F. Beek, H. Crezee, and T. M. van Gulik. “Cryoablation induces greater inflammatory and coagulative responses than radiofrequency ablation or laser induced thermotherapy in a rat liver model”. In: *Surgery* 147.5 (2010), pp. 686–695.

[45] J. M. Llovet, C. Brú, and J. Bruix. “Prognosis of hepatocellular carcinoma: the BCLC staging classification”. In: *Seminars in liver disease*. Vol. 19. 03. © 1999 by Thieme Medical Publishers, Inc. 1999, pp. 329–338.

[46] L. Solbiati, M. Ahmed, L. Cova, T. Ierace, M. Brioschi, and S. N. Goldberg. “Small liver colorectal metastases treated with percutaneous radiofrequency ablation: local response rate and long-term survival with up to 10-year follow-up”. In: *Radiology* 265.3 (2012), pp. 958–968.

[47] Y.-s. Kim, H. K. Lim, H. Rhim, M. W. Lee, D. Choi, W. J. Lee, et al. “Ten-year outcomes of percutaneous radiofrequency ablation as first-line therapy of early hepatocellular carcinoma: analysis of prognostic factors”. In: *Journal of hepatology* 58.1 (2013), pp. 89–97.

[48] I. Sucandy, S. Cheek, B. J. Golas, A. Tsung, D. A. Geller, and J. W. Marsh. “Longterm survival outcomes of patients undergoing treatment with radiofrequency ablation for hepatocellular carcinoma and metastatic colorectal cancer liver tumors”. In: *HPB* 18.9 (2016), pp. 756–763.

[49] M. Ahmed, L. Solbiati, C. L. Brace, D. J. Breen, M. R. Callstrom, J. W. Charboneau, et al. “Image-guided tumor ablation: standardization of terminology and reporting criteria—a 10-year update”. In: *Journal of Vascular and Interventional Radiology* 25.11 (2014), pp. 1691–1705.

[50] S. Clasen and P. L. Pereira. “Magnetic resonance guidance for radiofrequency ablation of liver tumors”. In: *Journal of Magnetic Resonance Imaging* 27.2 (2008), pp. 421–433.

[51] J. Du, H.-L. Li, B. Zhai, S. Chang, and F.-H. Li. “Radiofrequency ablation for hepatocellular carcinoma: utility of conventional ultrasound and contrast-enhanced ultrasound in guiding and assessing early therapeutic response and short-term follow-up results”. In: *Ultrasound in medicine & biology* 41.9 (2015), pp. 2400–2411.

[52] Y. H. Koh, J.-I. Choi, H. B. Kim, and M. J. Kim. “Computed tomographic-guided radiofrequency ablation of recurrent or residual hepatocellular carcinomas around retained iodized oil after transarterial chemoembolization”. In: *Korean journal of radiology* 14.5 (2013), pp. 733–742.

[53] L.-H. Lee, J.-I. Hwang, Y.-C. Cheng, C.-Y. Wu, S.-W. Lee, S.-S. Yang, et al. “Comparable outcomes of ultrasound versus computed tomography in the guidance of radiofrequency ablation for hepatocellular carcinoma”. In: *PloS one* 12.1 (2017), e0169655.

[54] J.-E. Kim, Y.-s. Kim, H. Rhim, H. K. Lim, M. W. Lee, D. Choi, S. W. Shin, and S. K. Cho. “Outcomes of patients with hepatocellular carcinoma referred for percutaneous radiofrequency ablation at a tertiary center: analysis focused on the feasibility with the use of ultrasonography guidance”. In: *European journal of radiology* 79.2 (2011), e80–e84.

[55] M. W. Lee, H. Rhim, D. I. Cha, Y. J. Kim, D. Choi, Y.-s. Kim, and H. K. Lim. “Percutaneous radiofrequency ablation of hepatocellular carcinoma: fusion imaging guidance for management of lesions with poor conspicuity at conventional sonography”. In: *American Journal of Roentgenology* 198.6 (2012), pp. 1438–1444.

[56] J. Stattaus, H. Kuehl, S. Ladd, T. Schroeder, G. Antoch, H. A. Baba, J. Barkhausen, and M. Forsting. “CT-guided biopsy of small liver lesions: visibility, artifacts, and corresponding diagnostic accuracy”. In: *Cardiovascular and interventional radiology* 30.5 (2007), pp. 928–935.

[57] S. Clasen, A. Boss, D. Schmidt, J. Fritz, C. Schraml, C. Claussen, and P. Pereira. “Magnetic resonance imaging for hepatic radiofrequency ablation”. In: *European journal of radiology* 59.2 (2006), pp. 140–148.

[58] F. K. Wacker and B. Hensen. "Clinical Evaluation of MR-guided Microwave Ablation and Thermometry of Primary and Secondary Liver Malignancies". Hannover Medical School, Germany. 2016.

[59] S. Roujol, M. Ries, B. Quesson, C. Moonen, and B. D. de Senneville. "Real-time MR-thermometry and dosimetry for interventional guidance on abdominal organs". In: *Magnetic Resonance in Medicine* 63.4 (2010), pp. 1080–1087.

[60] U. Kägebein, O. Speck, F. Wacker, and B. Hensen. "Motion correction in proton resonance frequency–based thermometry in the liver". In: *Topics in Magnetic Resonance Imaging* 27.1 (2018), pp. 53–61.

[61] C. W. König, P. L. Pereira, J. Trübenbach, J. Fritz, S. H. Duda, F. Schick, and C. D. Claussen. "MR Imaging–Guided Adrenal Biopsy Using an Open Low-Field-Strength Scanner and MR Fluoroscopy". In: *American Journal of Roentgenology* 180.6 (2003), pp. 1567–1570.

[62] J. Fritz, N. Tzaribachev, C. Thomas, J. Carrino, C. Claussen, J. Lewin, and P. Pereira. "Evaluation of MR imaging guided steroid injection of the sacroiliac joints for the treatment of children with refractory enthesitis-related arthritis". In: *European radiology* 21.5 (2011), pp. 1050–1057.

[63] J. S. Lewin, S. G. Nour, and J. L. Duerk. "Magnetic resonance image-guided biopsy and aspiration". In: *Topics in Magnetic Resonance Imaging* 11.3 (2000), pp. 173–183.

[64] J. Fritz, C. Thomas, S. Clasen, C. D. Claussen, J. S. Lewin, and P. L. Pereira. "Freehand real-time MRI-guided lumbar spinal injection procedures at 1.5 T: feasibility, accuracy, and safety". In: *American Journal of Roentgenology* 192.4 (2009), W161–W167.

[65] J. W. Genant, J. E. Vandevenne, A. G. Bergman, C. F. Beaulieu, S. T. Kee, A. M. Norbash, and P. Lang. "Interventional musculoskeletal procedures performed by using MR imaging guidance with a vertically open MR unit: assessment of techniques and applicability". In: *Radiology* 223.1 (2002), pp. 127–136.

[66] J. Stattaus, S. Maderwald, M. Forsting, J. Barkhausen, and M. E. Ladd. "MR-guided core biopsy with MR fluoroscopy using a short, wide-bore 1.5-Tesla scanner: Feasibility and initial results". In: *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 27.5 (2008), pp. 1181–1187.

[67] B. C. Meyer, A. Brost, D. L. Kraitchman, W. D. Gilson, N. Strobel, J. Hornegger, J. S. Lewin, and F. K. Wacker. “Percutaneous punctures with MR imaging guidance: comparison between MR imaging–enhanced fluoroscopic guidance and real-time MR imaging guidance”. In: *Radiology* 266.3 (2013), pp. 912–919.

[68] A. H. Mahnken, A. Buecker, E. Spuentrup, G. A. Krombach, D. Henzler, R. W. Günther, and J. Tacke. “MR-guided radiofrequency ablation of hepatic malignancies at 1.5 T: initial results”. In: *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 19.3 (2004), pp. 342–348.

[69] N. V. Tsekos, A. Özcan, and E. Christoforou. “A prototype manipulator for magnetic resonance-guided interventions inside standard cylindrical magnetic resonance imaging scanners”. In: *Journal of biomechanical engineering* 127.6 (2005), pp. 972–980.

[70] E. Christoforou, E. Akbudak, A. Ozcan, M. Karanikolas, and N. V. Tsekos. “Performance of interventions with manipulator-driven real-time MR guidance: implementation and initial in vitro tests”. In: *Magnetic resonance imaging* 25.1 (2007), pp. 69–77.

[71] U. Kägebein, F. Godenschweger, B. S. Armstrong, G. Rose, F. K. Wacker, O. Speck, and B. Hensen. “Percutaneous MR-guided interventions using an optical Moiré Phase tracking system: Initial results”. In: *PloS one* 13.10 (2018), e0205394.

[72] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino. “Augmented reality: A class of displays on the reality-virtuality continuum”. In: *Telemanipulator and telepresence technologies*. Vol. 2351. International Society for Optics and Photonics. 1995, pp. 282–293.

[73] R. T. Azuma. “A survey of augmented reality”. In: *Presence: Teleoperators & Virtual Environments* 6.4 (1997), pp. 355–385.

[74] J. Brooker. “The polytechnic ghost: Pepper’s ghost, metempsychosis and the magic lantern at the royal polytechnic institution”. In: *Early Popular Visual Culture* 5.2 (2007), pp. 189–206.

[75] J. Rekimoto and K. Nagao. “The world through the computer: Computer augmented interaction with real world environments”. In: *Proceedings of the 8th annual ACM symposium on User interface and software technology*. ACM. 1995, pp. 29–36.

[76] M. Billinghurst, A. J. Clark, and G. A. Lee. “A Survey of Augmented Reality”. In: *Foundations and Trends in Human-computer Interaction* 8.2 (2015), pp. 73–272.

[77] O. Bimber and R. Raskar. *Spatial augmented reality: Merging real and virtual worlds*. Wellesley Mass.: A K Peters, 2005.

[78] C. Harrison, H. Benko, and A. D. Wilson. "OmniTouch". In: *Proceedings of the 24th annual ACM symposium on User interface software and technology - UIST '11*. Ed. by J. Pierce, M. Agrawala, and S. Klemmer. New York, New York, USA: ACM Press, 2011, p. 441.

[79] K. A. Gavaghan, M. Peterhans, T. Oliveira-Santos, and S. Weber. "A portable image overlay projection device for computer-aided open liver surgery". In: *IEEE Transactions on Biomedical Engineering* 58.6 (2011), pp. 1855–1864.

[80] O. Hilliges, D. Kim, S. Izadi, M. Weiss, and A. Wilson. "HoloDesk". In: *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*. Ed. by J. A. Konstan, E. H. Chi, and K. Höök. New York, New York, USA: ACM Press, 2012, p. 2421.

[81] W. Broll. "Augmentierte Realität". In: *Virtual und Augmented Reality (VR / AR): Grundlagen und Methoden der Virtuellen und Augmentierten Realität*. Ed. by R. Dörner, W. Broll, P. Grimm, and B. Jung. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 241–294.

[82] I. E. Sutherland. "A head-mounted three dimensional display". In: *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*. ACM. 1968, pp. 757–764.

[83] L. Qian, A. Barthel, A. Johnson, G. Osgood, P. Kazanzides, N. Navab, and B. Fuerst. "Comparison of optical see-through head-mounted displays for surgical interventions with object-anchored 2D-display". In: *International journal of computer assisted radiology and surgery* 12.6 (2017), pp. 901–910.

[84] E. Rae, A. Lasso, M. S. Holden, E. Morin, R. Levy, and G. Fichtinger. "Neurosurgical burr hole placement using the Microsoft HoloLens". In: *Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling*. Vol. 10576. 2018, 105760T.

[85] H. L. Pryor, T. A. Furness, and E. Viirre. "The Virtual Retinal Display: A new Display Technology using Scanned Laser Light". In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 42.22 (1998), pp. 1570–1574.

[86] C. Jang, K. Bang, S. Moon, J. Kim, S. Lee, and B. Lee. "Retinal 3D". In: *ACM Transactions on Graphics* 36.6 (2017), pp. 1–13.

[87] M. Bajura, H. Fuchs, and R. Ohbuchi. "Merging virtual objects with the real world: Seeing ultrasound imagery within the patient". In: *ACM SIGGRAPH Computer Graphics* 26.2 (1992), pp. 203–210.

[88] M. Das, F. Sauer, U. J. Schoepf, A. Khamene, S. K. Vogt, S. Schaller, et al. "Augmented Reality Visualization for CT-guided Interventions: System Description, Feasibility, and Initial Evaluation in an Abdominal Phantom 1". In: *Radiology* 240.1 (2006), pp. 230–235.

[89] Z. Zhang. "Perspective Camera". In: *Computer Vision: A Reference Guide*. Ed. by K. Ikeuchi. Boston, MA: Springer US, 2014, pp. 590–592.

[90] G. Bradski and A. Kaehler. *Learning OpenCV: Computer vision with the OpenCV library*. " O'Reilly Media, Inc.", 2008.

[91] V. Chari and A. Veeraraghavan. "Lens Distortion, Radial Distortion". In: *Computer Vision*. Springer, 2014, pp. 443–445.

[92] Z. Zhang. "A flexible new technique for camera calibration". In: *IEEE Transactions on pattern analysis and machine intelligence* 22 (2000).

[93] W. Burger. *Zhang's Camera Calibration Algorithm: In-Depth Tutorial and Implementation*. Tech. rep. HGB16-05. Hagenberg, Austria: University of Applied Sciences Upper Austria, School of Informatics, Communications and Media, Dept. of Digital Media, May 2016.

[94] R. Tsai. "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses". In: *IEEE Journal on Robotics and Automation* 3.4 (1987), pp. 323–344.

[95] J. Heikkila and O. Silven. "A four-step camera calibration procedure with implicit image correction". In: *cvpr*. Vol. 97. 1997, p. 1106.

[96] J. Weng, P. Cohen, and M. Herniou. "Camera calibration with distortion models and accuracy evaluation". In: *IEEE Transactions on Pattern Analysis & Machine Intelligence* 10 (1992), pp. 965–980.

[97] O. D. Faugeras, Q.-T. Luong, and S. J. Maybank. "Camera self-calibration: Theory and experiments". In: *European conference on computer vision*. Springer. 1992, pp. 321–334.

[98] G. Falcao, N. Hurtos, and J. Massich. "Plane-based calibration of a projector-camera system". In: *VIBOT master* 9.1 (2008), pp. 1–12.

[99] F. Sadlo, T. Weyrich, R. Peikert, and M. Gross. "A practical structured light acquisition system for point-based geometry and texture". In: *Point-based graphics, 2005*. Ed. by M. Alexa and D. Fellner. Aire-la-Ville: Eurographics Association, 2005, pp. 89–145.

[100] M. Kimura, M. Mochimaru, and T. Kanade. “Projector Calibration using Arbitrary Planes and Calibrated Camera”. In: *IEEE Conference on Computer Vision and Pattern Recognition, 2007*. Los Alamitos, Calif.: IEEE Computer Society, 2007, pp. 1–2.

[101] J. Liao and L. Cai. “A calibration method for uncoupling projector and camera of a structured light system”. In: *IEEE/ASME International Conference on Advanced Intelligent Mechatronics, 2008*. Piscataway, NJ: IEEE Service Center, 2008, pp. 770–774.

[102] K. Yamauchi, H. Saito, and Y. Sato. “Calibration of a structured light system by observing planar object from unknown viewpoints”. In: *19th International Conference on Pattern Recognition, 2008*. Piscataway, NJ: IEEE, 2008, pp. 1–4.

[103] W. GAO, L. WANG, and Z.-Y. HU. “Flexible Calibration of a Portable Structured Light System through Surface Plane”. In: *Acta Automatica Sinica* 34.11 (2008), pp. 1358–1362.

[104] D. Moreno and G. Taubin. “Simple, accurate, and robust projector-camera calibration”. In: *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*. IEEE. 2012, pp. 464–471.

[105] H. Anwar, I. Din, and K. Park. “Projector calibration for 3D scanning using virtual target images”. In: *International Journal of Precision Engineering and Manufacturing* 13.1 (2012), pp. 125–131.

[106] J. Drareni, S. Roy, and P. Sturm. “Geometric video projector auto-calibration”. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition workshops, 2009*. Piscataway, NJ: IEEE, 2009, pp. 39–46.

[107] J. Draréni, S. Roy, and P. Sturm. “Methods for geometrical video projector calibration”. In: *Machine Vision and Applications* 23.1 (2012), pp. 79–89.

[108] S. Audet and M. Okutomi. “A user-friendly method to geometrically calibrate projector-camera systems”. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition workshops, 2009*. Piscataway, NJ: IEEE, 2009, pp. 47–54.

[109] I. Martynov, J.-K. Kamarainen, and L. Lensu. “Projector Calibration by “Inverse Camera Calibration””. In: *Image Analysis*. Ed. by A. Heyden and F. Kahl. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 536–544.

[110] S.-Y. Park and G. G. Park. “Active Calibration of Camera-Projector Systems Based on Planar Homography”. In: *20th International Conference on Pattern Recognition (ICPR), 2010*. Piscataway, NJ: IEEE, 2010, pp. 320–323.

[111] S. K. Nayar, G. Krishnan, M. D. Grossberg, and R. Raskar. "Fast separation of direct and global components of a scene using high frequency illumination". In: *ACM Transactions on Graphics* 25.3 (2006), p. 935.

[112] Y. Xu and D. G. Aliaga. "Robust pixel classification for 3D modeling with structured light". In: *Proceedings of Graphics Interface 2007*. Ed. by C. G. Healey. New York, NY: ACM, 2007, p. 233.

[113] B. Preim and C. Botha. "Chapter 18 - Image-Guided Surgery and Augmented Reality". In: *Visual Computing for Medicine (Second Edition)*. Ed. by B. Preim and C. Botha. Second Edition. Boston: Morgan Kaufmann, 2014, pp. 625–663.

[114] S. Garrido-Jurado, R. Muñoz-Salinas, F. Madrid-Cuevas, and R. Medina-Carnicer. "Generation of fiducial marker dictionaries using mixed integer linear programming". In: *Pattern Recognition* 51 (2016), pp. 481–491.

[115] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez. "Automatic generation and detection of highly reliable fiducial markers under occlusion". In: *Pattern Recognition* 47.6 (2014), pp. 2280–2292.

[116] S. Garrido-Jurado, R. Muñoz-Salinas, F. Madrid-Cuevas, and M. Marín-Jiménez. "Automatic generation and detection of highly reliable fiducial markers under occlusion". In: *Pattern Recognition* 47.6 (2014), pp. 2280–2292.

[117] J. Maclaren, B. S. R. Armstrong, R. T. Barrows, K. A. Danishad, T. Ernst, C. L. Foster, et al. "Measurement and Correction of Microscopic Head Motion during Magnetic Resonance Imaging of the Brain". In: *PloS one* 8.6 (2012), e48088.

[118] G. Reitmayr and T. Drummond. "Going out: robust model-based tracking for outdoor augmented reality." In: *ISMAR*. Vol. 6. 2006, pp. 109–118.

[119] G. Lowe. "SIFT-The Scale Invariant Feature Transform". In: *Int. J.* 2 (2004), pp. 91–110.

[120] H. Bay, T. Tuytelaars, and L. Van Gool. "Surf: Speeded up robust features". In: *European conference on computer vision*. Springer. 2006, pp. 404–417.

[121] A. M. Franz, T. Haidegger, W. Birkfellner, K. Cleary, T. M. Peters, and L. Maier-Hein. "Electromagnetic tracking in medicine—a review of technology, validation, and applications". In: *IEEE transactions on medical imaging* 33.8 (2014), pp. 1702–1725.

[122] S. Feiner, B. MacIntyre, T. Höllerer, and A. Webster. “A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment”. In: *Personal Technologies* 1.4 (1997), pp. 208–217.

[123] F. Sauer, S. Vogt, and A. Khamene. “Augmented Reality”. In: *Image-Guided Interventions: Technology and Applications*. Ed. by T. Peters and K. Cleary. Boston, MA: Springer US, 2008, pp. 81–119.

[124] C. Bichlmeier, F. Wimmer, S. M. Heining, and N. Navab. “Contextual Anatomic Mimesis Hybrid In-Situ Visualization Method for Improving Multi-Sensory Depth Perception in Medical Augmented Reality”. In: *6th IEEE and ACM International Symposium on Mixed and Augmented Reality, 2007*. Piscataway, NJ: IEEE, 2007, pp. 1–10.

[125] E. Kruijff, J. E. Swan, and S. Feiner. “Perceptual issues in augmented reality revisited”. In: *9th IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2010*. Ed. by T. Höllerer. Piscataway, NJ: IEEE, 2010, pp. 3–12.

[126] J. Fischer, D. Bartz, and W. Straßer. “Occlusion handling for medical augmented reality using a volumetric phantom model”. In: *Proceedings of the ACM symposium on Virtual reality software and technology*. Ed. by R. W. H. Lau and G. Baciu. New York, NY: ACM, 2004, p. 174.

[127] C. Furmanski, R. Azuma, and M. Daily. “Augmented-reality visualizations guided by cognition: perceptual heuristics for combining visible and obscured information”. In: *International Symposium on Mixed and Augmented Reality*. Los Almitos, Calif: IEEE Computer Society, 2002, pp. 215–320.

[128] N. Haouchine, J. Dequidt, I. Peterlik, E. Kerrien, M.-O. Berger, and S. Cotin. “Image-guided simulation of heterogeneous tissue deformation for augmented reality during hepatic surgery”. In: *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. Piscataway, NJ: IEEE, 2013, pp. 199–208.

[129] B. Marques, N. Haouchine, R. Plantefeve, and S. Cotin. “Improving depth perception during surgical augmented reality”. In: *ACM SIGGRAPH 2015 Posters*. Ed. by Unknown. New York, NY: ACM, 2015, p. 1.

[130] T. Ropinski, F. Steinicke, and K. Hinrichs. “Visually Supporting Depth Perception in Angiography Imaging”. In: *Smart graphics*. Ed. by A. Butz. Vol. 4073. Lecture Notes in Computer Science. Berlin: Springer, 2006, pp. 93–104.

[131] M. Kersten-Oertel, S. J.-S. Chen, and D. L. Collins. "An evaluation of depth enhancing perceptual cues for vascular volume visualization in neurosurgery". In: *IEEE transactions on visualization and computer graphics* 20.3 (2014), pp. 391–403.

[132] T. Sielhorst, C. Bichlmeier, S. M. Heining, and N. Navab. "Depth Perception – A Major Issue in Medical AR: Evaluation Study by Twenty Surgeons". In: *Medical image computing and computer-assisted intervention - MICCAI 2006*. Ed. by R. Larsen. Vol. 4190. Lecture Notes in Computer Science. Berlin: Springer, 2006, pp. 364–372.

[133] I. Viola and E. Gröller. "Smart Visibility in Visualization". In: *Computational Aesthetics in Graphics, Visualization and Imaging*. Ed. by Laszlo Neumann, Mateu Sbert, Bruce Gooch, and Werner Purgathofer. The Eurographics Association, 2005.

[134] J. Kruger, J. Schneider, and R. Westermann. "ClearView: An interactive context preserving hotspot visualization technique". In: *IEEE transactions on visualization and computer graphics* 12.5 (2006), pp. 941–948.

[135] F. Ritter, C. Hansen, V. Dicken, O. Konrad, B. Preim, and H.-O. Peitgen. "Real-time illustration of vascular structures". In: *IEEE transactions on visualization and computer graphics* 12.5 (2006), pp. 877–884.

[136] C. Hansen, J. Wieferich, F. Ritter, C. Rieder, and H.-O. Peitgen. "Illustrative visualization of 3D planning models for augmented reality in liver surgery". In: *International journal of computer assisted radiology and surgery* 5.2 (2010), pp. 133–141.

[137] K. Lawonn, M. Luz, and C. Hansen. "Improving spatial perception of vascular models using supporting anchors and illustrative visualization". In: *Computers & Graphics* 63 (2017), pp. 37–49.

[138] B. T. Phong. "Illumination for computer generated pictures". In: *Communications of the ACM* 18.6 (1975), pp. 311–317.

[139] M. E. Mundel. *Motion And Time Study-Principles And Practice*. Prentice-Hall, Inc.; New Jersy, 2013.

[140] J. J. LaViola Jr, E. Kruijff, R. P. McMahan, D. Bowman, and I. P. Poupyrev. *3D user interfaces: theory and practice*. Addison-Wesley Professional, 2017.

[141] D. A. Bowman, E. Kruijff, J. J. LaViola Jr, and I. Poupyrev. *3D user interfaces: theory and practice*. Addison-Wesley, 2004.

[142] D. Wigdor and D. Wixon. *Brave NUI world: designing natural user interfaces for touch and gesture*. Elsevier, 2011.

[143] M. Turk and G. Robertson. “Perceptual user interfaces”. In: *Communications of the ACM* 43.3 (2000), pp. 32–32.

[144] A. Mulder. “Hand gestures for HCI”. In: *Hand Centered Studies of Human Movement Project, Technical Report* (1996), pp. 96–1.

[145] A. Kendon. “How gesture can become like words”. In: *Cross-Cultural Perspective in Nonverbal Communication* (1988).

[146] A. Mewes, B. Hensen, F. Wacker, and C. Hansen. “Touchless interaction with software in interventional radiology and surgery: a systematic literature review”. In: *International journal of computer assisted radiology and surgery* 12.2 (2017), pp. 291–305.

[147] A. Alapetite. “Speech recognition for the anaesthesia record during crisis scenarios”. In: *international journal of medical informatics* 77.7 (2008), pp. 448–460.

[148] A. Perrakis, W. Hohenberger, and T. Horbach. “Integrated operation systems and voice recognition in minimally invasive surgery: comparison of two systems”. In: *Surgical endoscopy* 27.2 (2013), pp. 575–579.

[149] G. P. Mylonas, K.-W. Kwok, A. Darzi, and G.-Z. Yang. “Gaze-contingent motor channelling and haptic constraints for minimally invasive robotic surgery”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008*. Springer, 2008, pp. 676–683.

[150] N. T. Clancy, G. P. Mylonas, G.-Z. Yang, and D. S. Elson. “Gaze-contingent autofocus system for robotic-assisted minimally invasive surgery”. In: *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*. IEEE. 2011, pp. 5396–5399.

[151] D. Stoyanov, G. P. Mylonas, and G.-Z. Yang. “Gaze-contingent 3d control for focused energy ablation in robotic assisted surgery”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008*. Springer, 2008, pp. 347–355.

[152] B. Hatscher, M. Luz, L. E. Nacke, N. Elkmann, V. Müller, and C. Hansen. “GazeTap: Towards Hands-free Interaction in the Operating Room”. In: *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. ICMI 2017. Glasgow, UK: ACM, 2017, pp. 243–251.

[153] B. Hatscher and C. Hansen. “Hand, Foot or Voice: Alternative Input Modalities for Touchless Interaction in the Medical Domain”. In: *Proceedings of the 2018 on International Conference on Multimodal Interaction*. ICMI ’18. Boulder, CO, USA: ACM, 2018, pp. 145–153.

[154] J. P. Wachs, H. I. Stern, Y. Edan, M. Gillam, J. Handler, C. Feied, and M. Smith. “A gesture-based tool for sterile browsing of radiology images”. In: *Journal of the American Medical Informatics Association* 15.3 (2008), pp. 321–323.

[155] C. Kirmizibayrak, N. Radeva, M. Wakid, J. Philbeck, J. Sibert, and J. Hahn. “Evaluation of gesture based interfaces for medical volume visualization tasks”. In: *Proceedings of the 10th international conference on Virtual reality continuum and its applications in industry*. ACM. 2011, pp. 69–74.

[156] L. C. Ebert, G. Hatch, G. Ampanozi, M. J. Thali, and S. Ross. “You can’t touch this touch-free navigation through radiological images”. In: *Surgical innovation* 19.3 (2012), pp. 301–307.

[157] G. C. S. Ruppert, L. O. Reis, P. H. J. Amorim, T. F. de Moraes, and J. V. L. da Silva. “Touchless gesture user interface for interactive image visualization in urological surgery”. In: *World journal of urology* 30.5 (2012), pp. 687–691.

[158] A. M. Hötker, M. B. Pitton, P. Mildenberger, and C. Düber. “Speech and motion control for interventional radiology: requirements and feasibility”. In: *International journal of computer assisted radiology and surgery* 8.6 (2013), pp. 997–1002.

[159] M. Strickland, J. Tremaine, G. Brigley, and C. Law. “Using a depth-sensing infrared camera system to access and manipulate medical imaging from within the sterile operating field”. In: *Canadian Journal of Surgery* 56.3 (2013), E1.

[160] J. H. Tan, C. Chao, M. Zawaideh, A. C. Roberts, and T. B. Kinney. “Informatics in radiology: Developing a touchless user interface for intraoperative image control during interventional radiology procedures”. In: *Radiographics* 33.2 (2013), E61–E70.

[161] Y. A. Yusoff, A. H. Basori, and F. Mohamed. “Interactive Hand and Arm Gesture Control for 2D Medical Image and 3D Volumetric Medical Visualization”. In: *Procedia-Social and Behavioral Sciences* 97 (2013), pp. 723–729.

[162] B. Kocev, F. Ritter, and L. Linsen. “Projector-based surgeon–computer interaction on deformable surfaces”. In: *International journal of computer assisted radiology and surgery* 9.2 (2014), pp. 301–312.

[163] C. Chao, J. Tan, E. M. Castillo, M. Zawaideh, A. C. Roberts, and T. B. Kinney. “Comparative efficacy of new interfaces for intra-procedural imaging review: the Microsoft Kinect, Hillcrest Labs Loop Pointer, and the Apple iPad”. In: *Journal of digital imaging* 27.4 (2014), pp. 463–469.

[164] M. G. Jacob and J. P. Wachs. “Context-based hand gesture recognition for the operating room”. In: *Pattern Recognition Letters* 36 (2014), pp. 196–203.

[165] É. S. Silva and M. A. F. Rodrigues. “Design and Evaluation of a Gesture-Controlled System for Interactive Manipulation of Medical Images and 3D Models”. In: *SBC Journal on Interactive Systems* 5.3 (2014), pp. 53–65.

[166] M. T. Nouei, A. V. Kamyad, A. R. Soroush, and S. Ghazalbash. “A comprehensive operating room information system using the Kinect sensors and RFID”. In: *Journal of clinical monitoring and computing* 29.2 (2015), pp. 251–261.

[167] R. Wipfli, V. Dubois-Ferrière, S. Budry, P. Hoffmeyer, and C. Lovis. “Gesture-Controlled Image Management for Operating Room: A Randomized Crossover Study to Compare Interaction Using Gestures, Mouse, and Third Person Relaying”. In: *PloS one* 11.4 (2016), e0153596.

[168] K. O’Hara, G. Gonzalez, A. Sellen, G. Penney, A. Varnavas, H. Mentis, et al. “Touchless interaction in surgery”. In: *Communications of the ACM* 57.1 (2014), pp. 70–77.

[169] S. Soutschek, J. Penne, J. Hornegger, and J. Kornhuber. “3-d gesture-based scene navigation in medical imaging applications using time-of-flight cameras”. In: *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW’08. IEEE Computer Society Conference on*. IEEE. 2008, pp. 1–6.

[170] D. S. Lopes, P. D. de Figueiredo Parreira, S. F. Paulo, V. Nunes, P. A. Rego, M. C. Neves, P. S. Rodrigues, and J. A. Jorge. “On the utility of 3D hand cursors to explore medical volume datasets with a touchless interface”. In: *Journal of Biomedical Informatics* 72 (2017), pp. 140–149.

[171] F. M. Sánchez-Margallo, J. A. Sánchez-Margallo, J. L. Moyano-Cuevas, E. M. Pérez, and J. Maestre. “Use of natural user interfaces for image navigation during laparoscopic surgery: initial experience”. In: *Minimally Invasive Therapy & Allied Technologies* 26.5 (2017), pp. 253–261.

[172] P. Hughes, N. Nestorov, N. Healy, N. Sheehy, and N. O’Hare. “Comparing the utility and usability of the Microsoft Kinect and Leap Motion sensor devices in the context of their application for gesture control of biomedical images”. In: European Congress of Radiology 2015. 2015.

[173] N. Bizzotto, A. Costanzo, L. Bizzotto, D. Regis, A. Sandri, and B. Magnan. “Leap Motion Gesture Control With OsiriX in the Operating Room to Control Imaging First Experiences During Live Surgery”. In: *Surgical innovation* (2014), pp. 655–656.

[174] L. Ebert, P. Flach, M. Thali, and S. Ross. “Out of touch–A plugin for controlling OsiriX with gestures using the leap controller”. In: *Journal of Forensic Radiology and Imaging* 2.3 (2014), pp. 126–128.

[175] S. Mauser and O. Burgert. “Touch-free, gesture-based control of medical devices and software based on the leap motion controller”. In: *Stud Health Technol Inform* 196 (2014), pp. 265–270.

[176] G. M. Rosa and M. L. Elizondo. “Use of a gesture user interface as a touchless image navigation system in dental surgery: Case series report”. In: *Imaging science in dentistry* 44.2 (2014), pp. 155–160.

[177] J. Pauchot, L. Di Tommaso, A. Lounis, M. Benassarou, P. Mathieu, D. Bernot, and S. Aubry. “Leap Motion Gesture Control With Carestream Software in the Operating Room to Control Imaging Installation Guide and Discussion”. In: *Surgical innovation* (2015), pp. 615–620.

[178] A. Mewes, P. Saalfeld, O. Riabikin, M. Skalej, and C. Hansen. “A gesture-controlled projection display for CT-guided interventions”. In: *International journal of computer assisted radiology and surgery* 11.1 (2016), pp. 157–164.

[179] P. Saalfeld, A. Mewes, M. Luz, B. Preim, and C. Hansen. “Comparative Evaluation of Gesture and Touch Input for Medical Software”. In: De Gruyter Oldenbourg, 2015.

[180] A. Opromolla, V. Volpi, A. Ingrosso, S. Fabri, C. Rapuano, D. Passalacqua, and C. M. Medaglia. “A Usability Study of a Gesture Recognition System Applied During the Surgical Procedures”. In: *Design, User Experience, and Usability: Interactive Experience Design*. Springer, 2015, pp. 682–692.

[181] B. J. Park, T. Jang, J. W. Choi, and N. Kim. “Gesture-controlled interface for contactless control of various computer programs with a hooking-based keyboard and mouse-mapping technique in the operating room”. In: *Computational and Mathematical Methods in Medicine* 2016 (2016).

[182] T. Kipshagen, M. Graw, V. Tronnier, M. Bonsanto, and U. Hofmann. “Touch- and marker-free interaction with medical software”. In: *World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany*. Springer. 2009, pp. 75–78.

[183] L. A. Schwarz, A. Bigdelou, and N. Navab. “Learning gestures for customizable human-computer interaction in the operating room”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011*. Springer, 2011, pp. 129–136.

[184] A. Bigdelou, L. Schwarz, and N. Navab. “An adaptive solution for intra-operative gesture-based human-machine interaction”. In: *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*. ACM. 2012, pp. 75–84.

[185] S. Jalaliniya, J. Smith, M. Sousa, L. Büthe, and T. Pederson. “Touch-less interaction with medical images using hand & foot gestures”. In: *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*. ACM. 2013, pp. 1265–1274.

[186] J. Hettig, A. Mewes, O. Riabikin, M. Skalej, B. Preim, and C. Hansen. “Exploration of 3D Medical Image Data for Interventional Radiology using Myoelectric Gesture Control”. In: *Proceedings of Eurographics Workshop on Visual Computing for Biology and Medicine*. The Eurographics Association, 2015, pp. 177–185.

[187] T. Baudel and M. Beaudouin-Lafon. “Charade: remote control of objects using free-hand gestures”. In: *Commun. ACM* 36.7 (1993), pp. 28–35.

[188] B. Suelze, R. Agten, P. B. Bertrand, T. Vandenryt, R. Thoelen, P. Vandervoort, and L. Grieten. “Waving at the Heart: Implementation of a Kinect-based real-time interactive control system for viewing cineangiogram loops during cardiac catheterization procedures”. In: *Computing in Cardiology Conference (CinC), 2013*. IEEE. 2013, pp. 229–232.

[189] H. M. Mentis, K. O’Hara, G. Gonzalez, A. Sellen, R. Corish, A. Criminisi, R. Trivedi, and P. Theodore. “Voice or Gesture in the Operating Room”. In: *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM. 2015, pp. 773–780.

[190] D. Stevenson, H. Gardner, W. Neilson, E. Beenen, S. Gananadha, J. Fergusson, et al. “Evidence from the surgeons: gesture control of image data displayed during surgery”. In: *Behaviour & Information Technology* 35.12 (2016), pp. 1063–1079.

[191] A. Seitel, L. Maier-Hein, S. Schawo, B. A. Radeleff, S. A. Mueller, F. Pianka, et al. “In-vitro evaluation of different visualization approaches for computer assisted targeting in soft tissue”. In: *International journal of computer assisted radiology and surgery* 2 (2007), S188–S190.

[192] T. Oliveira-Santos, B. Klaeser, T. Weitzel, T. Krause, L.-P. Nolte, M. Peterhans, and S. Weber. “A navigation system for percutaneous needle interventions based on PET/CT images: design, workflow and error analysis of soft tissue and bone punctures”. In: *Computer aided surgery : official journal of the International Society for Computer Aided Surgery* 16.5 (2011), pp. 203–219.

[193] A. Seitel, N. Bellemann, M. Hafezi, A. M. Franz, M. Servatius, A. Saffari, et al. “Towards markerless navigation for percutaneous needle insertions”. In: *International journal of computer assisted radiology and surgery* 11.1 (2016), pp. 107–117.

[194] U. Kägebein, F. Godenschweger, A. Boese, and O. Speck. “Automatic Image Plane Alignment for Percutaneous MR-guided Interventions Using an Optical Moiré Phase Tracking System”. In: *In Magnetic Resonance Materials in Physics, Biology and Medicin*. Vol. volume 28.2015. Edinburgh, 2015, pp. 159–160.

[195] B. Carl, M. Bopp, B. Saß, B. Voellger, and C. Nimsky. “Implementation of augmented reality support in spine surgery”. In: *European Spine Journal* (2019).

[196] S. Nicolau, L. Soler, D. Mutter, and J. Marescaux. “Augmented reality in laparoscopic surgical oncology”. In: *Surgical oncology* 20.3 (2011), pp. 189–201.

[197] B. Le Roy, E. Ozgur, B. Koo, E. Buc, and A. Bartoli. “Augmented reality guidance in laparoscopic hepatectomy with deformable semi-automatic computed tomography alignment (with video)”. In: *Journal of visceral surgery* (2019).

[198] Y. Kim, H. Kim, and Y. O. Kim. “Virtual reality and augmented reality in plastic surgery: a review”. In: *Archives of plastic surgery* 44.3 (2017), p. 179.

[199] L. Besharati Tabrizi and M. Mahvash. “Augmented reality–guided neurosurgery: accuracy and intraoperative application of an image projection technique”. In: *Journal of neurosurgery* 123.1 (2015), pp. 206–211.

[200] A. Meola, F. Cutolo, M. Carbone, F. Cagnazzo, M. Ferrari, and V. Ferrari. “Augmented reality in neurosurgery: a systematic review”. In: *Neurosurgical review* 40.4 (2017), pp. 537–548.

[201] T. Sielhorst, M. Feuerstein, and N. Navab. “Advanced medical displays: A literature review of augmented reality”. In: *Journal of Display Technology* 4.4 (2008), pp. 451–467.

[202] A. Dey, M. Billinghurst, R. W. Lindeman, and J. Swan. “A systematic review of 10 years of augmented reality usability studies: 2005 to 2014”. In: *Frontiers in Robotics and AI* 5 (2018), p. 37.

[203] F. K. Wacker, S. Vogt, A. Khamene, J. A. Jesberger, S. G. Nour, D. R. Elgort, et al. “An Augmented Reality System for MR Image–guided Needle Biopsy: Initial Results in a Swine Model 1”. In: *Radiology* 238.2 (2006), pp. 497–504.

[204] F. Sauer, A. Khamene, and S. Vogt. "An augmented reality navigation system with a single-camera tracker: System design and needle biopsy phantom trial". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2002, pp. 116–124.

[205] C. R. Weiss, D. R. Marker, G. S. Fischer, G. Fichtinger, A. J. Machado, and J. A. Carrino. "Augmented reality visualization using image-overlay for MR-guided interventions: system description, feasibility, and initial evaluation in a spine phantom". In: *American Journal of Roentgenology* 196.3 (2011), W305–W307.

[206] G. Fichtinger, A. Deguet, K. Masamune, E. Balogh, G. S. Fischer, H. Mathieu, et al. "Image overlay guidance for needle insertion in CT scanner". In: *IEEE transactions on biomedical engineering* 52.8 (2005), pp. 1415–1424.

[207] P. U-Thainual, J. Fritz, C. Moonjaita, T. Ungi, A. Flammang, J. A. Carrino, G. Fichtinger, and I. Iordachita. "MR image overlay guidance: system evaluation for preclinical use". In: *International journal of computer assisted radiology and surgery* 8.3 (2013), pp. 365–378.

[208] M. Sugimoto, H. Yasuda, K. Koda, M. Suzuki, M. Yamazaki, T. Tezuka, et al. "Image overlay navigation by markerless surface registration in gastrointestinal, hepatobiliary and pancreatic surgery". In: *Journal of hepato-biliary-pancreatic sciences* 17.5 (2010), pp. 629–636.

[209] K. A. Gavaghan, S. Anderegg, M. Peterhans, T. Oliveira-Santos, and S. Weber. "Augmented reality image overlay projection for image guided open liver ablation of metastatic liver cancer". In: *Workshop on Augmented Environments for Computer-Assisted Interventions*. Springer. 2011, pp. 36–46.

[210] K. Gavaghan, T. Oliveira-Santos, M. Peterhans, M. Reyes, H. Kim, S. Anderegg, and S. Weber. "Evaluation of a portable image overlay projector for the visualisation of surgical navigation data: phantom studies". In: *International journal of computer assisted radiology and surgery* 7.4 (2012), pp. 547–556.

[211] R. Krempien, H. Hoppe, L. Kahrs, S. Daeuber, O. Schorr, G. Eggers, et al. "Projector-based augmented reality for intuitive intraoperative guidance in image-guided 3D interstitial brachytherapy". In: *International journal of radiation oncology, biology, physics* 70.3 (2008), pp. 944–952.

[212] D. Black, J. Hettig, M. Luz, C. Hansen, R. Kikinis, and H. Hahn. “Auditory feedback to support image-guided medical needle placement”. In: *International journal of computer assisted radiology and surgery* 12.9 (2017), pp. 1655–1663.

[213] L. T. De Paolis and V. de Luca. “Augmented visualization with depth perception cues to improve the surgeon’s performance in minimally invasive surgery”. In: *Medical & biological engineering & computing* (2018).

[214] F. Bork, B. Fuers, A.-K. Schneider, F. Pinto, C. Graumann, and N. Navab. “Auditory and Visio-Temporal Distance Coding for 3-Dimensional Perception in Medical Augmented Reality”. In: *2015 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE. 2015, pp. 7–12.

[215] *MRIsafety.com*. English. http://www.mrisafety.com.

[216] Z. C. Marton, R. B. Rusu, and M. Beetz. “On fast surface reconstruction methods for large and noisy point clouds”. In: *2009 IEEE international conference on robotics and automation*. IEEE. 2009, pp. 3218–3223.

[217] M. Minge, M. Thüring, I. Wagner, and C. V. Kuhr. “The meCUE questionnaire: a modular tool for measuring user experience”. In: *Advances in Ergonomics Modeling, Usability & Special Populations*. Springer, 2017, pp. 115–128.

[218] S. G. Hart. “NASA-task load index (NASA-TLX); 20 years later”. In: *Proceedings of the human factors and ergonomics society annual meeting*. Vol. 50. 9. Sage publications Sage CA: Los Angeles, CA. 2006, pp. 904–908.

[219] S. G. Hart and L. E. Staveland. “Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research”. In: *Advances in psychology*. Vol. 52. Elsevier, 1988, pp. 139–183.

[220] Y.-L. Park, S. Elayaperumal, B. Daniel, S. C. Ryu, M. Shin, J. Savall, et al. “Real-Time Estimation of 3-D Needle Shape and Deflection for MRI-Guided Interventions”. In: *IEEE/ASME transactions on mechatronics : a joint publication of the IEEE Industrial Electronics Society and the ASME Dynamic Systems and Control Division* 15.6 (2010), pp. 906–915.

[221] F. Heinrich, G. Schmidt, K. Bornemann, A. L. Roethe, W. I. Essayed, and C. Hansen. “Visualization concepts to improve spatial perception for instrument navigation in image-guided surgery”. In: *Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling*. Vol. 10951. International Society for Optics and Photonics. 2019, p. 1095125.

[222] E. Pannicke, U. Kägebein, and R. Vick. “Interventional Coil Design”. In: *Proceedings of the 12th Interventional MRI Symposium*. Boston, USA, 2018.

[223] J. Westwood. “3D Physics-Based Registration of 2D Dynamic MRI Data”. In: *Medicine Meets Virtual Reality 22: NextMed/MMVR22* 220 (2016), p. 432.

[224] R. Xu and G. A. Wright. “GPU accelerated dynamic respiratory motion model correction for MRI-guided cardiac interventions”. In: *Computer methods and programs in biomedicine* 136 (2016), pp. 31–43.

[225] G. Narita, Y. Watanabe, and M. Ishikawa. “Dynamic projection mapping onto deforming non-rigid surface using deformable dot cluster marker”. In: *IEEE transactions on visualization and computer graphics* 23.3 (2016), pp. 1235–1248.

[226] J. Alpers, C. Hansen, K. Ringe, and C. Rieder. “CT-Based Navigation Guidance for Liver Tumor Ablation”. In: *Eurographics Workshop on Visual Computing for Biology and Medicine*. The Eurographics Association, 2017.

[227] A. Mewes, F. Heinrich, U. Kägebein, B. Hensen, F. Wacker, and C. Hansen. “Projector-based augmented reality system for interventional visualization inside MRI scanners”. In: *The International Journal of Medical Robotics and Computer Assisted Surgery* 15.1 (2019), e1950.

[228] A. Mewes, F. Heinrich, B. Hensen, F. Wacker, K. Lawonn, and C. Hansen. “Concepts for augmented reality visualisation to support needle guidance inside the MRI”. In: *Healthcare technology letters* 5.5 (2018), pp. 172–176.

[229] W. A. Rutala, M. S. White, M. F. Gergen, and D. J. Weber. “Bacterial contamination of keyboards: efficacy and functional impact of disinfectants”. In: *Infection Control & Hospital Epidemiology* 27.4 (2006), pp. 372–377.

[230] M. Schultz, J. Gill, S. Zubairi, R. Huber, and F. Gordin. “Bacterial contamination of computer keyboards in a teaching hospital”. In: *Infection Control & Hospital Epidemiology* 24.4 (2003), pp. 302–303.

[231] A. Hübler, C. Hansen, O. Beuing, M. Skalej, and B. Preim. “Workflow Analysis for Interventional Neuroradiology using Frequent Pattern Mining.” In: *Proceedings of the Annual Meeting of the German Society of Computer- and Robot-Assisted Surgery*. 2014, pp. 165–168.

[232] G. El-Shallaly, B. Mohammed, M. Muhtaseb, A. Hamouda, and A. Nassar. “Voice recognition interfaces (VRI) optimize the utilization of theatre staff and time during laparoscopic cholecystectomy”. In: *Minimally Invasive Therapy & Allied Technologies* 14.6 (2005), pp. 369–371.

[233] I. A. Salama and S. D. Schwaitzberg. “Utility of a voice-activated system in minimally invasive surgery”. In: *Journal of Laparoendoscopic & Advanced Surgical Techniques* 15.5 (2005), pp. 443–446.

[234] C.-A. O. Nathan, V. Chakradeo, K. Malhotra, H. D’Agostino, and R. Patwardhan. “The voice-controlled robotic assist scope holder AESOP for the endoscopic approach to the sella”. In: *Skull Base* 16.3 (2006), p. 123.

[235] Y.-T. Li, M. Jacob, G. Akingba, and J. P. Wachs. “A cyber-physical management system for delivering and monitoring surgical instruments in the OR”. In: *Surgical innovation* 20.4 (2013), pp. 377–384.

[236] A. Alapetite. “Impact of noise and other factors on speech recognition in anaesthesia”. In: *International journal of medical informatics* 77.1 (2008), pp. 68–77.

[237] M. Visentini-Scarzanella, G. P. Mylonas, D. Stoyanov, and G.-Z. Yang. “i-brush: A gaze-contingent virtual paintbrush for dense 3d reconstruction in robotic assisted surgery”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009*. Springer, 2009, pp. 353–360.

[238] R. Wen, W.-L. Tay, B. P. Nguyen, C.-B. Chng, and C.-K. Chui. “Hand gesture guided robot-assisted surgery based on a direct augmented reality interface”. In: *Computer methods and programs in biomedicine* 116.2 (2014), pp. 68–80.

[239] F. Hartmann and A. Schlaefer. “Feasibility of touch-less control of operating room lights”. In: *International journal of computer assisted radiology and surgery* 8.2 (2013), pp. 259–268.

[240] R. H. Gong, Ö. Güler, M. Kürklüoglu, J. Lovejoy, and Z. Yaniv. “Interactive initialization of 2D/3D rigid registration”. In: *Medical physics* 40.12 (2013), p. 121911.

[241] R. Reilink, G. De Bruin, M. Franken, M. Mariani, S. Misra, and S. Stramigioli. “Endoscopic camera control by head movements for thoracic surgery”. In: *Biomedical Robotics and Biomechatronics (BioRob), 2010 3rd IEEE RAS and EMBS International Conference on*. IEEE. 2010, pp. 510–515.

[242] A. Nishikawa, T. Hosoi, K. Koara, D. Negoro, A. Hikita, S. Asano, et al. “FAce MOUSe: A novel human-machine interface for controlling the position of a laparoscope”. In: *Robotics and Automation, IEEE Transactions on* 19.5 (2003), pp. 825–841.

[243] B. Hatscher, M. Luz, and C. Hansen. “Foot Interaction Concepts to Support Radiological Interventions”. In: *Mensch und Computer 2017 - Tagungsband*. Ed. by M. Burghardt, R. Wimmer, C. Wolff, and C. Womser-Hacker. Regensburg: Gesellschaft für Informatik e.V., 2017, pp. 93–104.

[244] M. Riduwan, A. H. Basori, and F. Mohamed. “Finger-based Gestural Interaction for Exploration of 3D Heart Visualization”. In: *Procedia-Social and Behavioral Sciences* 97 (2013), pp. 684–690.

[245] L. Gallo. “A study on the degrees of freedom in touchless interaction”. In: *SIGGRAPH Asia 2013 Technical Briefs*. ACM. 2013, p. 28.

[246] F. Weichert, D. Bachmann, B. Rudak, and D. Fisseler. “Analysis of the accuracy and robustness of the leap motion controller”. In: *Sensors* 13.5 (2013), pp. 6380–6393.

[247] J. Guna, G. Jakus, M. Pogačnik, S. Tomažič, and J. Sodnik. “An analysis of the precision and reliability of the leap motion sensor and its suitability for static and dynamic tracking”. In: *Sensors* 14.2 (2014), pp. 3702–3720.

[248] M. Nielsen, M. Störring, T. B. Moeslund, and E. Granum. “A procedure for developing intuitive and ergonomic gesture interfaces for HCI”. In: *International gesture workshop*. Springer. 2003, pp. 409–420.

[249] K. A. Ericsson and H. A. Simon. *Protocol analysis: Verbal reports as data.* 1984.

[250] J. Prümper. “Software-evaluation based upon ISO 9241 Part 10”. In: *Vienna Conference on Human Computer Interaction*. Springer. 1993, pp. 255–265.

[251] C. P. Gerba, A. L. Wuollet, P. Raisanen, and G. U. Lopez. “Bacterial contamination of computer touch screens”. In: *American journal of infection control* 44.3 (2016), pp. 358–360.

[252] M. J. Riffe, S. R. Yutzy, Y. Jiang, M. D. Twieg, C. J. Blumenthal, D. P. Hsu, et al. “Device localization and dynamic scan plane selection using a wireless magnetic resonance imaging detector array”. In: *Magnetic resonance in medicine* 71.6 (2014), pp. 2243–2249.

[253] M. T. Mazilu, A. Z. Faranesh, J. A. Derbyshire, R. J. Lederman, and M. S. Hansen. “Low-cost MRI compatible interface device for interactive scan plane control”. In: *Proceedings of the 19th Annual Meeting of ISMRM, Montreal, Canada*. 2011, p. 3752.

[254] F. Güttler. *Gesture control: The hand for the contactless control of the MRI during interventional procedures*. 2018.

[255] M. A. Rube, A. B. Holbrook, B. F. Cox, R. Buciuc, and A. Melzer. “Wireless mobile technology to improve workflow and feasibility of MR-guided percutaneous interventions”. In: *International journal of computer assisted radiology and surgery* 10.5 (2015), pp. 665–676.

[256] E. Pannicke, B. Hatscher, B. Hensen, A. Mewes, C. Hansen, F. Wacker, and R. Vick. “MR Compatible and Sterile Gesture Interaction for Interventions”. In: *Proceedings of the 12th Interventional MRI Symposium*. 2018.

[257] J. Brooke. “SUS-A quick and dirty usability scale”. In: *Usability evaluation in industry* 189.194 (1996), pp. 4–7.

[258] A. Bangor, P. Kortum, and J. Miller. “Determining what individual SUS scores mean: Adding an adjective rating scale”. In: *Journal of usability studies* 4.3 (2009), pp. 114–123.

[259] J. Hettig, P. Saalfeld, M. Luz, M. Becker, M. Skalej, and C. Hansen. “Comparison of gesture and conventional interaction techniques for interventional neuroradiology”. In: *International journal of computer assisted radiology and surgery* 12.9 (2017), pp. 1643–1653.

[260] S. Cronin and G. Doherty. “Touchless computer interfaces in hospitals: A review”. In: *Health informatics journal* (2018), p. 1460458217748342.

[261] B. Hatscher, A. Mewes, E. Pannicke, U. Kägebein, F. Wacker, C. Hansen, and B. Hensen. “Touchless scanner control to support MRI-guided interventions”. In: *International Journal of Computer Assisted Radiology and Surgery* (Sept. 2019).

Publication List

[1] A. Mewes, P. Saalfeld, O. Riabikin, M. Skalej, and C. Hansen. “A gesture-controlled projection display for CT-guided interventions”. In: *International journal of computer assisted radiology and surgery* 11.1 (2016), pp. 157–164.

[2] A. Mewes, B. Hensen, F. Wacker, and C. Hansen. “Touchless interaction with software in interventional radiology and surgery: a systematic literature review”. In: *International journal of computer assisted radiology and surgery* 12.2 (2017), pp. 291–305.

[3] A. Mewes, F. Heinrich, B. Hensen, F. Wacker, K. Lawonn, and C. Hansen. “Concepts for augmented reality visualisation to support needle guidance inside the MRI”. In: *Healthcare technology letters* 5.5 (2018), pp. 172–176.

[4] A. Mewes, F. Heinrich, U. Kägebein, B. Hensen, F. Wacker, and C. Hansen. “Projector-based augmented reality system for interventional visualization inside MRI scanners”. In: *The International Journal of Medical Robotics and Computer Assisted Surgery* 15.1 (2019), e1950.

[5] P. Saalfeld, A. Mewes, M. Luz, B. Preim, and C. Hansen. “Comparative Evaluation of Gesture and Touch Input for Medical Software”. In: De Gruyter Oldenbourg, 2015.

[6] P. Saalfeld, A. Mewes, C. Hansen, and B. Preim. “Gaze-Based Annotations: Labels on Demand”. In: *Proceedings of the Annual Meeting of the German Society of Computer- and Robot-Assisted Surgery*. Bremen, 2015, pp. 261–266.

[7] J. Hettig, A. Mewes, O. Riabikin, M. Skalej, B. Preim, and C. Hansen. “Exploration of 3D Medical Image Data for Interventional Radiology using Myoelectric Gesture Control”. In: *Proceedings of Eurographics Workshop on Visual Computing for Biology and Medicine*. The Eurographics Association, 2015, pp. 177–185.

[8] S. Weiß, A. Schnurr, A. Mewes, T. Hoffmann, D. Schindele, M. Schostak, and C. Hansen. “Mobile Augmented Reality and 3D Printing to Involve Patients in Treatment Decisions for Prostate Cancer”. In: *Proceedings of the Annual Meeting of the German Society of Computer- and Robot-Assisted Surgery (CURAC)*. Bern, Schweiz, 2016, pp. 11–16.

[9] E. Pannicke, B. Hatscher, B. Hensen, A. Mewes, C. Hansen, F. Wacker, and R. Vick. “MR Compatible and Sterile Gesture Interaction for Interventions”. In: *Proceedings of the 12th Interventional MRI Symposium*. 2018.

[10] B. Hatscher, A. Mewes, E. Pannicke, U. Kägebein, F. Wacker, C. Hansen, and B. Hensen. “Touchless scanner control to support MRI-guided interventions”. In: *International Journal of Computer Assisted Radiology and Surgery* (Sept. 2019).

List of Acronyms

AR augmented reality
BCLC Barcelona Clinic Liver Cancer
BMBF German Federal Ministry of Education and Research
CT computed tomography
DOF degree of freedom
EASL European Association for the Study of the Liver
ERDF European Regional Development Fund
FPS frames per second
FOV field of view
GUI graphical user interface
HCC hepatocellular carcinoma
HCI human-computer interaction
HIFU high-intensity focused ultrasound
HMD head-mounted display
HUD head-up display
IRE irreversible electroporation
iMRI interventional MRI
LMC Leap Motion Controller
MP Moiré phase
MPT Moiré phase tracking
MR magnetic resonance
MRI magnetic resonance imaging
MWA microwave ablation
NUI natural user interface

OR	operating room
RFA	radiofrequency ablation
SRC	Scanner Remote Control
SUS	System Usability Scale
UI	user interface
US	ultrasound
VR	virtual reality

List of Formulas

L_d	direct light component
L_g	global light component
L^+	total light on lit surface patch
L^-	total light on unlit surface patch
α	fraction of activated projector pixels
β	fraction of brightness of an activated pixel emitted by a deactivated element
P	pixel intensity interval
p	pixel intensity value
P_{on}	pixel intensity interval for activated pixels
P_{off}	pixel intensity interval for deactivated pixels
$\hat{H}$	local homography
R	rotation matrix
t	translation vector

List of Figures

2.1 The adapted BCLC staging and treatment strategy according to the EASL guidelines. 6
2.2 Zones of hyperthermic ablation and its effects on them. 8
2.3 Workflow of an MRI-guided intervention. 11
2.4 Workflow of an MRI-guided percutaneous thermal ablation. 12
2.5 Finding the planned entry point with a finger on the patient's skin in real-time images from the MRI. 13
2.6 Schematic drawing of the real-time needle guidance with the MPT system. 15

3.1 Reality-Virtuality Continuum. 17
3.2 Different display categories used to create AR: head-attached, handheld, and spatial. 19
3.3 Sutherland's AR system with an optical see-through HMD and mechanical tracking in 1968. 21
3.4 The Microsoft HoloLens optical see-through display. 22
3.5 Schematic view of a retinal display forming laser beams to be scanned directly to the user's eyes. 22
3.6 HMD with attached camera to show virtual US images aligned with the real patient. 23
3.7 Projector-based AR. Left: in marketing; right: in arts. 25
3.8 Perspective projection with the pinhole camera model. The image plane is put in front of the centre of projection to facilitate calculations. 26
3.9 Intrinsic and extrinsic parameters of the pinhole camera. 27
3.10 Projector calibration after Falcao et al. 30
3.11 Pixel classification scenarios. 33
3.12 Calculation of local homographies between the projected and the captured pattern image. 35
3.13 Three ArUco markers in different sizes (5,6 and 8 bits). 38
3.14 Robustness of ArUco marker boards against occlusion [0]. 39
3.15 Moiré marker and moiré pattern used for pose tracking. ©metriainnovation.com 2019 40

3.16 Passive infrared light optical marker tracking. 41

3.17 Combination of a wireframe and opaque blood vessels on a laparoscopic video feed. 46

3.18 Tumour and blood vessels during a laparoscopic liver resection. Left: virtual elements seem to be in front of the liver; right: semi-transparency by alpha blending improves depth perception. 46

3.19 The focus and context visualisation forms a virtual window. When moving, the perspective changes and allows for better spatial perception through motion parallax. 47

3.20 Illustrative rendering of vessels to improve depth perception. 49

3.21 Comparison between volume rendering and illustrative visualisation of vessels on the liver in an AR scene. 50

3.22 Comparison of projections of volume rendered and illustrative visualisation onto a liver. 51

3.23 Different modalities to recognise touchless input during interventions. 53

3.24 A physician performs arm gestures in front of a screen and the Microsoft Kinect. 55

3.25 Schematic drawing of a physician performing hand gestures above a stereo camera. 55

3.26 A physician interacting with the interventional software via the Myo armband. 56

3.27 Three zones in the image to activate gesture recognition, perform the gesture, and deactivate again. 58

4.1 A video see-through AR HMD is used to overlay the patient with 2D images used by Wacker et al. 61

4.2 An optical see-through mirror is placed in front of the MRI bore to present 2D MR images aligned with a phantom. 62

4.3 Projection of a tumour model onto the brain. 63

4.4 The handheld projector AR navigation proposed by Gavaghan et al. 64

4.5 Crosshair navigation aid with depth-indicating progress bar and virtual needle to support needle punctures. 64

4.6 Projected explicit navigation cues proposed by Krempien et al. 65

4.7 Schematic drawing of the complete hardware setup. ©2018 John Wiley and Sons 67

4.8 Hardware setup in the MRI scanner room. 69

4.9 One of the projected grey code patterns on the chessboard during projector calibration from the view of the ceiling-mounted wide-angle camera. ©2018 John Wiley and Sons 70

4.10 Estimation of the fixed ChArUco marker pose for registration of the MRI coordinate system with the camera coordinate system. ©2018 John Wiley and Sons 71
4.11 Calibration body filled with candle wax for MRI imaging and with a ChArUco board attached to estimate the phantoms pose in camera coordinates. ©2018 John Wiley and Sons 72
4.12 The different transforms involved in the registration process. 73
4.13 Projection of a planned insertion point and contours of a rib on the irregular surface of an abdomen phantom viewed from the feet side in the correct viewing position. ©2018 John Wiley and Sons 74
4.14 Visualisation of two-dimensional explicit navigation aids while moving the needle to the entry point, aligning, and inserting it. 76
4.15 Navigation visualisation to support needle navigation. 77
4.16 Exemplary view over the user's shoulder of the AR needle navigation inside the MRI bore. A combination of both *2D* and *3D* concept is shown. ©2018 IEEE 78
4.17 View of the projected scene from above with the wide-angle camera. ©2018 John Wiley and Sons 80
4.18 Measured task duration, entry point error and target distance error for each concept grouped by experience. 81
4.19 Nasa TLX results for each concept. 82
4.20 A flexible MRI coil integrated in a sterile drape for interventional use. 86

5.1 The angiography suite provides little space for additional displays. Therefore, the radiation shield on the left of the radiologist is used as a projection surface. 95
5.2 Prototype of interactive radiation shield. The GUI is back-projected onto the radiation shield. The gesture sensor is placed below. ©2015 CARS 96
5.3 The GUI of the image viewer. 97
5.4 The gesture set. 98
5.5 Users' valuations of the single gestures. 103
5.6 Usability and ergonomic scores from the ISONORM 9241/10 questionnaire. 104
5.7 Projector and LMC mounted onto the radiation shield with a custom 3D-printed mount. 105
5.8 Two radiologists in the scanner room (seen through a shielded window). 107
5.9 Survey results on "How do you currently control the MRI scanner?". 110
5.10 Survey results on usage of the functions before and during interventions. 110

5.11 Survey results on interventional usefulness of certain functions. 111
5.12 Hand gesture set for touchless MRI control. 112
5.13 GUI for direct, touchless MRI control. 113
5.14 Study setup. 114
5.15 Task duration of the six tasks. 117
5.16 Subjective workload score from the NASA-TLX questionnaire. 117
5.17 System usability derived from SUS questionnaires. 118

List of Tables

4.1 Results of the accuracy and duration measurements grouped by experience and concept. 81
4.2 meCUE scores for both visualisation concepts. 83

5.1 Durations of the gestures used in the subtasks of all tasks. 101
5.2 Pros and cons of the UI based on the comments of the users. 102

Questionnaire to Evaluate the Gesture Set (Chapter 5)

Hinweis: Die Auswertung der Studie und des folgenden Fragebogens erfolgt komplett anonym.

Fragebogen

Alter:

Geschlecht:	☐ weiblich	☐ männlich
Händigkeit:	☐ Linkshänder	☐ Rechtshänder
Sind Sie Medizinstudent/in?	☐ ja	☐ nein
Sind Sie Ärztin/Arzt?	☐ ja	☐ nein

Falls Sie Ärztin oder Arzt sind:

☐ Assistenzärztin/arzt ☐ Oberärztin/arzt ☐ Chefärztin/arzt

Welche Fachrichtung?:

Wie viele Jahre Berufserfahrung haben Sie:

Haben Sie Erfahrung mit:

Medizinischer Software	☐ ja	☐ nein
Betrachtung von Schichtbildern	☐ ja	☐ nein
Betrachtung von 3D-Modellen	☐ ja	☐ nein
Freihand-Gesten-Interaktion	☐ ja	☐ nein

Bewertung	**--**	**-**	**0**	**+**	**++**
***Greifgeste* zur 3D-Rotation**					
Die Benutzung der Geste ist einfach					
Das Ausführen der Geste fühlt sich natürlich an					
Das Ausführen der Geste ist *nicht* ermüdend					
Die Geste ist einfach zu merken					
Wie weit die Hand zur Ausführung der Geste geschlossen werden musst, ist verständlich					
An der Geste hat mir gefallen:					
An der Geste hat mir nicht gefallen:					

ID

Bewertung	--	-	0	+	++
***Wischgeste* zur schrittweisen Rotation**					
Die Benutzung der Geste ist einfach					
Das Ausführen der Geste fühlt sich natürlich an					
Das Ausführen der Geste ist *nicht* ermüdend					
Die Geste ist einfach zu merken					
Der Grad der Rotation entspricht meinen Vorstellungen					
An der Geste hat mir gefallen:					
An der Geste hat mir nicht gefallen:					

Bewertung	**--**	**-**	**0**	**+**	**++**
***Faustgeste* zum Zoomen und Verschieben**					
Die Benutzung der Geste ist einfach					
Das Ausführen der Geste fühlt sich natürlich an					
Das Ausführen der Geste ist *nicht* ermüdend					
Die Geste ist einfach zu merken					
Die Verschiebedistanz entspricht meinen Vorstellungen					

An der Geste hat mir gefallen:	
An der Geste hat mir nicht gefallen:	

Bewertung	**--**	**-**	**0**	**+**	**++**
***Kreiselgeste* zum Durchschalten der Slices**					
Die Benutzung der Geste ist einfach					
Das Ausführen der Geste fühlt sich natürlich an					
Das Ausführen der Geste ist *nicht* ermüdend					
Die Geste ist einfach zu merken					
Der Kreisradius bildet die Geschwindigkeit des Durchschaltens gut ab					

An der Geste hat mir gefallen:	
An der Geste hat mir nicht gefallen:	

Bewertung	**--**	**-**	**0**	**+**	**++**
***Zeigegeste* zum Selektieren**					
Die Benutzung der Geste ist einfach					
Das Ausführen der Geste fühlt sich natürlich an					
Das Ausführen der Geste ist *nicht* ermüdend					
Die Geste ist einfach zu merken					
An der Geste hat mir gefallen:					
An der Geste hat mir nicht gefallen:					

Bewertung	**--**	**-**	**0**	**+**	**++**
***Offene-Handgeste* als Anfangsgeste**					
Die Benutzung der Geste ist einfach					
Das Ausführen der Geste fühlt sich natürlich an					
Das Ausführen der Geste ist *nicht* ermüdend					
Die Geste ist einfach zu merken					
An der Geste hat mir gefallen:					
An der Geste hat mir nicht gefallen:					

Bewertung der Software

	Die Software ...	- -	-	o	+	++	Die Software ...
aa2	erfordert überflüssige Eingaben.	O	O	O	O	O	erfordert keine überflüssigen Eingaben.
sb1	liefert in unzureichendem Maße Informationen darüber, welche Eingaben zulässig oder nötig sind.	O	O	O	O	O	liefert in zureichendem Maße Informationen darüber, welche Eingaben zulässig oder nötig sind.
ek1	erschwert die Orientierung durch eine uneinheitliche Gestaltung.	O	O	O	O	O	erleichtert die Orientierung durch eine einheitliche Gestaltung.
lf1	erfordert viel Zeit zum Erlernen.	O	O	O	O	O	erfordert wenig Zeit zum Erlernen.
lf2	erfordert, dass man sich viele Details merken muss.	O	O	O	O	O	erfordert nicht, dass man sich viele Details merken muss.
lf3	ist schlecht ohne fremde Hilfe oder Handbuch erlernbar.	O	O	O	O	O	ist gut ohne fremde Hilfe oder Handbuch erlernbar.

	Die Software ...	--	-	o	+	++	Die Software ...
sk1	erzwingt eine unnötig starre Einhaltung von Bearbeitungsschritten.	○	○	○	○	○	erzwingt keine unnötig starre Einhaltung von Bearbeitungsschritten.
ft2	erfordert bei Fehlern im Großen und Ganzen einen hohen Korrekturaufwand.	○	○	○	○	○	erfordert bei Fehlern im Großen und Ganzen einen geringen Korrekturaufwand.

Falls Sie Ärztin oder Arzt sind:

Trifft nicht zu — Trifft zu

Bewertung	**--**	**-**	**0**	**+**	**++**
Ich könnte mir vorstellen diese Form der Gesteninteraktion im Operationsaal einzusetzen					

Kommentare

____

____

____

____