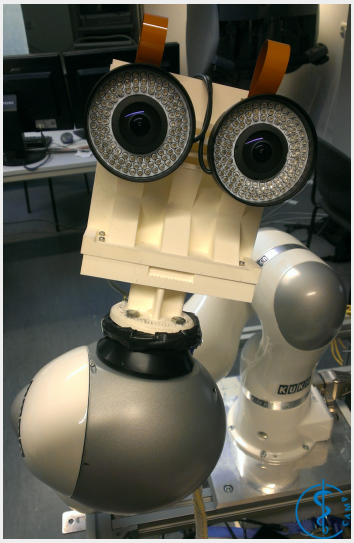
Computer Aided Medical Procedures
Prof. Dr. Nassir Navab



Dissertation

Seamless Intraoperative Robotic Imaging

Marco Esposito

Fakultät für Informatik
Technische Universität München

Technische Universität München
Fakultät für Informatik
Lehrstuhl für Informatikanwendungen in der Medizin

Seamless Intraoperative Robotic Imaging

Marco Esposito

Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende(r): Prof. Dr.-Ing. Darius Burschka

Prüfer der Dissertation: 1. Prof. Dr. Nassir Navab

2. Prof. Dr. Lena Maier-Hein

Die Dissertation wurde am 13.12.2018 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 26.04.2019 angenommen.

Marco Esposito
Seamless Intraoperative Robotic Imaging

Technische Universität München
Fakultät für Informatik
Lehrstuhl für Informatikanwendungen in der Medizin
Boltzmannstraße 3
85748 and Garching bei München

Abstract

Medical imaging has assumed a major role in modern medicine. The ability to precisely ascertain the cause and origin of the patient's condition before starting surgical treatment has removed a large part of the guesswork and imprecision that were previously involved. But also during the surgery, the adoption of imaging technologies has helped reducing the invasivity of the procedures and improving the short- and long-term clinical outcome.
However, the introduction of more and more tools into the operating room has also contributed to increase the complexity of modern surgery. Today, becoming a surgeon also involves learning to use the available technologies effectively and cope with their idiosyncrasies, while keeping focused on the actual objective: healing the patient. An optimal trade-off in this regard can be challenging to achieve, and the clinical outcome can be negatively affected by the cognitive burden and stress posed on the physician.

Such scenarios can occur when the surgeon needs combined anatomical and functional information to perform a surgery with success. Most intraoperative imaging modalities take the form of hand-held devices, each providing its own output image according to its position in space. The physician must hold both devices at the same time in order to scan the same region of space, and mentally orient and merge their output in order to identify structures of interest. One or both devices must then be put aside before carrying out the actual task, based on the acquired knowledge. We explore how collaborative robotic imaging can simplify this picture, by holding one of the devices for the physician and working in concert with them while producing one single, real-time multimodal image which is ready for visualization and can be provided at all times.

An additional obstacle for medical personnel can be represented by fickle equipment that poses many constraints on its handling. Freehand imaging techniques are subject to the limitations of currently available tracking technologies, such as *line-of-sight* requirements or the degradation of accuracy across the workspace or due to environmental factors. The operator is then required to prevent these problematics by constraining their own actions while pursuing the clinical objective. Failure to do so can compromise the system's performance or force to repeat the image acquisition altogether. We investigate the possibility to mitigate these issues by employing mathematical tools to partially overcome the precision limitations of electromagnetic tracking and make freehand 3D ultrasound more robust and reliable to use.

Zusammenfassung

Die medizinische Bildgebung hat in der modernen Medizin eine wichtige Rolle eingenommen. Die Fähigkeit, die Ursache und den Ursprung des Zustands des Patienten vor Beginn der chirurgischen Behandlung exakt zu bestimmen, hat einen großen Teil der zuvor bestehenden Fehleinschätzungen und Ungenauigkeiten beseitigt. Aber auch während der Operation hat die Einführung von Bildgebungstechnologien dazu beigetragen, die Invasivität der Eingriffe zu reduzieren und das kurz- sowie langfristige klinische Ergebnis zu verbessern.
Gleichzeitig hat auch die Benutzung von vielen verschiedenen Werkzeugen in den Operationssaal dazu beigetragen, die Komplexität der modernen Chirurgie zu erhöhen. Chirurg zu sein bedeutet heute auch, die verfügbaren Technologien effektiv zu nutzen und mit ihren Eigenheiten umzugehen, während man sich auf das eigentliche Hauptziel konzentriert: die Behandlung des Patienten. Ein optimaler Kompromiss in dieser Hinsicht kann schwierig zu erreichen sein, und das klinische Ergebnis kann durch die kognitive Belastung und den Stress für den Arzt negativ beeinflusst werden.

Solche Szenarien können auftreten, wenn der Chirurg kombinierte anatomische und funktionelle Bildinformationen benötigt, um eine Operation erfolgreich durchführen zu können. Die meisten intraoperativen Bildgebungsverfahren werden in Form von Handheld-Geräten durchgeführt, die jeweils ein individuelles Ausgabebild entsprechend ihrer Position im Raum liefern. Der Arzt muss beide Geräte gleichzeitig in den Händen halten, um den gleichen Raumbereich zu scannen, und die Ergebnisse mental orientieren und zusammenführen, um interessante Strukturen zu identifizieren. Ein oder beide Geräte müssen dann gelegt werden, um den eigentlichen chirurgischen Eingriff auf Grundlage der Bilddaten ausführen zu können. Wir untersuchen, wie die kollaborative Roboterbildgebung dies vereinfachen kann, indem ein robotischer Arm eines der Geräte für den Arzt hält und mit ihm zusammenarbeitet, während ein einziges, multimodales Echtzeit-Bild erzeugt wird, das visualisierungsbereit und jederzeit verfügbar ist.

Ein zusätzliches Hindernis für das medizinische Personal kann das unbeständige Equipment darstellen, das viele Einschränkungen bei der Handhabung mit sich bringt. Freihand-Bildgebungstechniken unterliegen den Grenzen der derzeit verfügbaren Tracking-Technologien, wie z.B. den Anforderungen an die Sichtlinie oder der Verschlechterung der Genauigkeit im gesamten Arbeitsbereich oder aufgrund von Umweltfaktoren. Der Anwender muss dann diese Probleme verhindern, indem er sein eigenes Handeln einschränkt und gleichzeitig das klinische Ziel verfolgt. Andernfalls kann die Leistung des Systems beeinträchtigt werden oder zur Notwendigkeit führen, die Bildaufnahme vollständig zu wiederholen. Wir untersuchen die Möglichkeit, diese Probleme durch den Einsatz mathematischer Werkzeuge zu entschärfen, um die Präzisionsgrenzen der elektromagnetischen Verfolgung teilweise zu überwinden und den Freihand-3D-Ultraschall robuster und zuverlässiger zu machen.

Acknowledegments

The purpose of this verbose and official document is to (only partially) summarize my professional activity at TUM and CAMP. Luckily I can take advantage of this page to add a more personal note, and to thank the wonderful people that I met during my experience there. I am extremely grateful to all of them, not only for their professional help, but also for the influence that they had on my personal development during this time.

First I would like to thank Prof. Nassir Navab for the opportunity that was given me. When I joined the chair I was only an overenthusiastic engineer. With patience, Nassir taught me to think on a different layer of abstraction, and to focus on the important things without getting lost in details. This skill is worth one thousand notions, and I will always cherish it. I would also like to thank Prof. Lena Maier-Hein for kindly accepting to be my second examiner and participating in this important step of my life, and Martina Hilla for her titanic efforts to maintain me in my blissful ignorance of what happened behind the scenes of the administration.

Then I would like to mention some senior figures that guided me throughout my career at the university. José Gardiazabal deserves a special mention here. As supervisor of my Master's thesis, we spent numerous nights playing with radioactive material, binding female body parts to a bed with black duct tape and danced around robots. This was enough to convince him to help me getting a position at the chair on the right side of the Atlantic Ocean. By that time, I was not new to the chair: Prof. Pascal Fallavollita had already built his own Italian gang at NARVIS, where I had a blast drilling bones and taking X-ray shots. I cannot thank him enough for introducing me to research, and Roberto Londei for accompanying me in that journey. But then it was time to move on to the IFL, where Philipp Matthies and Asli Okur helped Benjamin Frisch in taking care of me in my baby steps. It was only thanks to them and the rest of the IFL team if I had such a successful start. So I want to thank Julia Rackerseder for her work in the timezone opposite to mine; Benjamin Busam for giving eyes to the robot; Salvatore Virga for giving it a soul and a pet pig; Oliver Zettinig for his consulence on technical matters and on Austrian terminology; Rüdiger Göbl for the myriad of problems he found a solution to. The same goes for the rest of the team: Beatrice Lentes, Javier Esteban, Walter Simson, and many others. A special mention is due to Christoph Hennersperger and Maximilian Baust, who offered me their precious guidance during my hardest times in the lab. I also want to thank all the Garching residents, which are too many mention: Prof. Tobias Lasser, Christian Schulte zu Berge, Mehrdad Salehi, Iro Laina, Christian Rupprecht, ...

This list would not be complete without mentioning the group that left Pisa with me and moved to Munich before it was cool: Nicholas Kirk, Fausto Milletarì and Endri Koci. It was only with their support that I was able to start an adventure that brought me many times around the world, and to meet an endless list of awesome people, which I have to make unfairly short: Rebecca Rittstieg, Dimitar Kurtev, Stefan Nosović, Mario Garcia, Macarena Velasco Hodgson, Chiara Amat di San Filippo, Sarthak Pati, Navketan Verma. But among the Indian gang, Hemal Naik deserves a special mention for dragging me onto the CAMP rollercoaster first. Thank you, for this and all that followed.

I would like to dedicate this work to my parents, Carla Gianella and Mauro Esposito, for their unconditional support that they constantly provided me in all possible forms. And to Nicola, who with her unbounded talents and cheerfulness can make the sun shine on the rainiest day.

Contents

Part I

Introduction

1 Introduction

> *Your scientists were so preoccupied with whether they could that they didn't stop to think if they should.*
>
> **— Ian Malcolm**
> (Jurassic Park, M. Chrichton)

Among the profound and radical changes that took place in the 20th century, one of the most pervasive innovations can arguably be found in the introduction of electronics and informatics. The beginning of the new millennium brought constant performance improvements conjugated to a progressive miniaturization, to the point that a contemporary wristwatch can fit many times over the computational power that brought the man on the moon. While the first computers were considered extravagant toys that would never find large diffusion, soon they crept into every factory, office, house and pocket. One of the reasons is that they constitute the only technology able to manipulate the most important asset: information. A movie, a book, the blueprint of a house, a recipe, our schedule for tomorrow, the CAD model of a mechanical component, the current stock exchange rates, a medical scan can all be virtualized out of their physical support and transferred at the speed of light, automatically manipulated according to an algorithm of choice or concurrently shared among people located around the globe. It is hard to find a human activity which does not involve information, and hence cannot benefit of the "computer revolution". This change has been so disruptive that it has been recognized as one of the industrial revolutions.

The medical field has not been indifferent to this development, and a wave of research already turned into a proliferation of new tools available to physicians and surgeons. Most 3D imaging modalities could not be viable without the availability of abundant computational power, to the point that some are named after it (e.g. "Computed Tomography"). The digitalization of medical data unlocked the doors to sophisticated interventional planning and assistance, in the form of software and dedicated devices.

The application of electronics and informatics to medicine spawned entirely new research and industrial fields, with a high level of interdisciplinarity. It is an unfortunate case and the worry of several authors [89, 144] that the collaboration between personnel coming from an engineering and from a medical background is not yet optimal. While engineers tend to "fall in love" with their technology and focus exclusively on what their research would make possible, they often fail to recognize if such a thing would be needed or beneficial at all from a medical point of view. On the other hand, physicians and surgeons rarely have the time to properly evaluate new approaches that would present opportunities for a disruptive improvement in patient care or the safety of the medical personnel; the consequence is that they tend to prefer small and incremental changes that do not disrupt the workflows that required years of training, partly because of their very suboptimality in the first place.

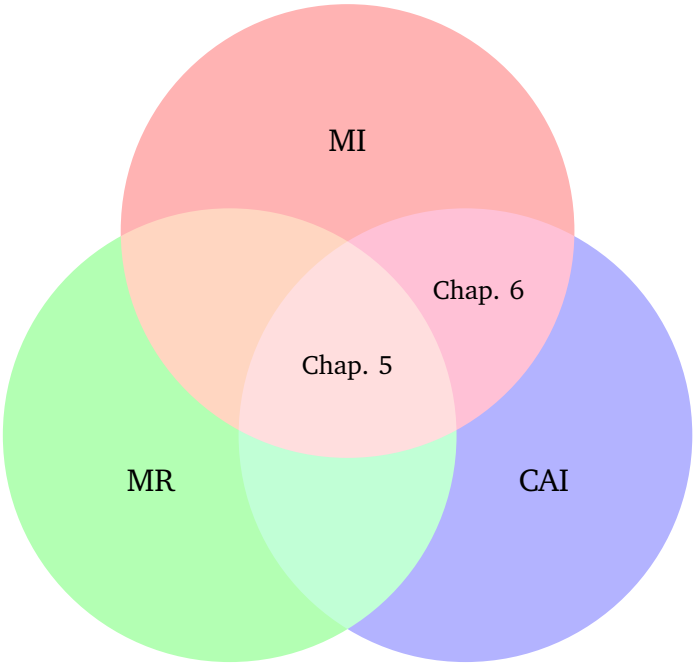


Fig. 1.1. Our contributions lie at the intersection of different established fields: Medical Imaging (MI), Computer Assisted Interventions (CAI), and Medical Robotics (MR). In Chapter 5 we will present a multimodal imaging device based on a collaborative robotic paradigm to provide intraoperative guidance. In Chapter 6 we will propose a method to improve the usability and the outcome of an imaging modality in intraoperative setups.

In response to the need of bridging the communities together, a growing number of institutions are bringing the researchers closer to the hospitals. Technical labs are being opened within medical facilities, as well as new centers for translational medicine, where early-adopting surgeons can evaluate and help develop the newest concepts. The NARVIS and IFL labs of the Chair for Computer Aided Medical Procedures (CAMP) of the Technical University of Munich belong to the first category. NARVIS hosts mostly research about cone beam CT, augmented reality and user interaction, while IFL focuses on ultrasound imaging, image registration and collaborative robotics. All these technologies are object of intensive research, for different and complementary reasons. Cone beam CT allows for high quality intraoperative imaging, but at the cost of danger to the patient and the surgeon through ionizing radiation; on the other hand, ultrasound imaging is fast and harmless but offers a lower quality due to the low dimensionality, the high density of artifacts and the limited tissue penetration. Augmented reality and User Interaction can be used to provide visualization and guidance with the ever growing amount of information available in the operating room, while collaborative robotics can help manage the increasing amount of devices and information by concurrently performing multimodal imaging without explicit control, or keeping a safety distance between the surgeon's instrument and sensitive regions via virtual fixtures.

This dissertation describes two innovative approaches that we developed in the IFL: both attempt at tackling the challenges of today's medical practice while avoiding to disrupt established workflows. Our aim is to **reduce complexity** with the introduction of our solutions into the clinical procedures compared to the existing state of the art.
This requires an unusual approach to the problem, which is reflected in the structure of this

text: rather than *vertically* deepening a specific topic, we will zoom out to consider the big picture of the technological means available to support the surgeon's work. We will discuss current imaging modalities and robotic devices, with their advantages as well as their obstacles to acceptance and usage. Finally we will present our proposed solutions and motivate them in this context.

The rest of **Part I** will consist in a brief overview of the history of surgery and of the resulting recurrent clinical setups that emerged. This will provide a context for motivating the constraints and acceptance factors that medical technologies are subject to. **Part II** will present the fields at the intersection of which this research lies: Medical Imaging (**Chapter 2**), Computer Assisted Interventions (CAI, **Chapter 3**) and Medical Robotics (**Chapter 4**).
Part III will go into the detail of our contributions.

In **Chapter 5** we will present an intraoperative imaging system for multimodal guidance during Sentinel Lymph Node (SLN) biopsy in the axilla for breast cancer staging. The current clinical routine involves open surgery to perform SLN dissection under visual guidance, supported by 1D gamma probes. By introducing a collaborative robotic assistant holding a gamma camera, we allow the surgeon to perform a less invasive needle biopsy under guidance of concurrent 2D ultrasound and gamma imaging. In contrast to other approaches, our system provides real-time imaging which is robust to tissue deformations in the area of the breast. Furthermore, the surgeon's workflow is not different than a standard ultrasound-guided needle biopsy: the introduction of the robot into the operating theater actually alleviates the burden posed on the physician, which must no longer actively handle the gamma camera to obtain the augmentation of the anatomical image with functional information. Our experiments on phantoms show the potential significant impact on the clinical outcome of the procedure.

Chapter 6 will be about the application of a Riemannian regularization framework to the tracking information used in freehand 3D ultrasound compounding. Most commercial ultrasound systems include an electromagnetic (EM) tracking system to follow the position of the probe in space and allow for the generation of a 3D volume from 2D slices acquired over time. This is due to the lack of a *line-of-sight* requirement, which is an important drawback of optical tracking systems. Unfortunately, EM systems are usually less accurate than optical devices. This is due in part to the effect of metallic objects in the neighborhood, which can distort the magnetic field generated by the system to locate the sensor in space; most of the literature about EM tracking error correction focuses on compensating for this distortion, via calibration approaches or by including redundant sensors. Instead, we will turn our attention on *jitter*, which can result in artifacts that make the image harder to interpret for the user. We will adopt a data processing method that, in contrast to common approaches, includes a term explicitely penalizing the deviation of the result from the original input trajectory. We will evaluate the effect of the application of this method on the volumes reconstructed by the same ultrasound images, according to the data originating from different tracking systems. We will also employ a strategy to directly measure the *local* quality of the compounded volumes, in contrast to *global* measures used to quantify the impact of distortion compensation approaches (such as Point Reconstruction Error).

Part IV will contain our final remarks and future outlook.

Fig. 1.2. **Depiction of a 19th century surgical theater.** The people in white are the surgical team, performing a mastectomy on a patient lying on the bed. The surrounding audience is composed of medical students, assisting to the surgery in the context of the 1889 medical class of the University of Pennsylvania. *The Agnew Clinic, Thomas Eakins, 1889. Image under Public Domain.*

1.1 History and Evolution of Surgery

The first examples of surgical interventions are prehistorical, as indicated by skull drilling and suture marks which can be identified in some archaeological remains. However, these signs testimony rather the prowess of a few pioneers than a systematic pursue of advancement in knowledge and praxis, which could only take place with a rigorous scientific approach in the modern Era [100]. The investigation of human anatomy, started during the Renaissance, prepared ground for a scientific and empirical study of surgery, which evolved during the Age of Enlightenment. The invention of **anesthesia** allowed to investigate more complex procedures [127], and the discovery of microscopic pathogens lead to a better understanding of how to minimize the risk of infection.

By the second half of the 18th century, several interventions had become possible. However, the largest part took the form of *open surgery*: the region of interest was exposed via incisions on the patient's skin. This revealed the interior of the patient not only to the surgeon's eyes and hands, but also to infecting agents. As a consequence, the wound took a long time to heal, not always without long term averse effects.

Arguably, the most important enabler of the ongoing surgical invasivity reduction can be found in the invention and advancements of **medical imaging**. While the "forerunners" of modern surgeons had to literally dig into the patient in order to find and address the source of the problem (as shown in Fig. 1.2), new technologies allow contemporary physicians to inspect the patient's body from the outside and offered a medium to store and share knowledge. The discovery of X-rays at the end of 19th century soon turned into a medical imaging application, spawning the entirely new and still vital field of **Radiology**.

To be fair, the long way to the development of **Minimal Invasive Surgery** (MIS) had already started by that time. While the record for the first organ to be inspected through reflected light can be traced back to the 11th century, in the study of the Arab physician Albukasim,

the first dedicated device was built only eight centuries later by the German physician Philipp Bozzini, which used it to inspect the human urethra. As in many other cases, his work found heavy resistance in the community, which readily defined his device "a toy" [178].

Luckily some of his colleagues did not follow the then popular motto "the larger the cut, the better the surgeon", and refined the technique in order to apply it to non-natural orifices for inspection (**Endoscopy**) and surgery: the first non-diagnostic **Laparoscopy** was performed in the early 1930s [178]. The biggest challenge was delivering light to the region of interest, in order to make it visible to the surgeon, without thermal issues related to the use of fire; the invention of the light bulb helped overcome this problem, and the development of flexible instruments led to a boom in the number of applications. The same principle applied to other regions of the human body beyond the abdominal cavity, such as the thoracic cavity (Video Assisted Thoracic Surgery, **VATS** [94]) and joint cavities (**Arthroscopy** [120]).

Given the adoption rate and the improvement to patient care that it enabled, laparoscopy can be counted amongst the most disruptive innovations in medicine, together with anesthesia and antibiotics [178]. Many procedures are now performed with a minimally invasive approach in more than 90% of cases, with significant improvement of post-operatory decourse [187].

Another important factor shaping today's medicine can be found in the proliferation of imaging modalities, and their combination resulting in **multimodal** imaging methods. The fusion of anatomical modalities (like X-ray based Computer Tomography - CT, or Magnetic Resonance Imaging - MRI), and functional imaging modalities (for example Single Photon Emission Computer Tomography - SPECT or Positronic Emission Tomography - PET) can combine the characteristics of both and allow to differentiate tissues that would look alike in a strictly anatomical image, facilitating the detection of tumoral masses and other anomalies. Interestingly, these possibilities were also not recognized right away: as in the case of the endoscope, the first SPECT/CT device was also judged to be "just a toy" [26].

The digitalization of medical imaging was a prerequisite for a much more systematic and integrated use of the produced images. Mechanical stereotaxy was integrated with **navigation** systems making use of optical, mechanical or electromagnetic tracking to localize the position of the instrument handled by the surgeon inside a pre- or intraoperative image, reducing the cognitive burden on the physician. Digital endoscopes, capable of capturing light outside of the human visible range, can visualize information that would otherwise be hidden; an example is the distribution of a contrast liquid injected into the patient, with the aim of locating vessels and avoid their damage during tissue resection.

A solution that recently found commercial viability and promises to deliver benefits to both patients and doctors is **robotic surgery** [56], which takes different forms. On one hand, many ad-hoc devices are being developed to enhance particular procedures in orthopedics, neurosurgery and many other fields; at the same time, teleoperated laparoscopy is finding its way into general surgery. Making use of an electronic interface and advanced 3D visualization, it offers a much better ergonomy over classic laparoscopy, possibly allowing the surgeon to work under reduced stress and minimizing the collateral damage when operating in cluttered parts of the body [28]. Several procedures have been found viable, and clinical studies are underway to demonstrate the clinical advantages for the patients in order to justify the

increased costs. The evolution of the technical solutions adopted, together with the integration of image-based and cognitive features, promises to provide a strong platform capable of further reducing the invasivity of the procedures by exploiting natural orifices and offering the surgeon smart, context-aware assistance.

These are only some of the milestones that were laid on the path to patient-personalized, minimally invasive, precision surgery. In the next three sections we will focus on each of the three aforementioned and complementary fields: Medical Imaging, Computer Aided Interventions and Medical Robotics. We do not aim at an exhaustive review, which would be unfeasible in the scope of this work; but rather to build a conceptual map of the state of the art and current research directions, in order to provide context and motivation to our contributions.

1.2 Common Clinical Setups

For an easier comprehension of the following sections, it is beneficiary to review the most common scenarios that can be encountered during surgical procedures. Along widely adopted open, laparoscopic, robotic and microsurgery, we will briefly introduce ultrasound-guided needle biopsy since it is of particular interest for the discussion of the presented contributions.

1.2.1 Open Surgery

General surgery is typically performed in a dedicated area of the hospital, called **Operating Room** (OR). The most distinctive trait of an OR is the particular care taken to ensure the sterility of the whole environment, achieved through the particular design of the room itself and furniture within, and through strict protocols applied to every person and object entering it.
The usual equipment (see Fig. 1.3) includes at least an **Operating Table**, that is a bed on which the patient can lie and be solidly fixated in order to prevent dangerous motions during the procedure. Many tables are adjustable in position and orientation to adapt to each procedure and phase of the surgery.
Surmounting the bed are placed **surgical lights**, usually implemented as a potent illumination system which can be oriented to keep the operating site clearly visible at all times. This can be challenging due to the large number of people, devices and tools involved.
At least a cart is in general necessary to hold the **surgical instruments**: scalpels, throngs, and more specialized tools. These must be completely sterile and are usually responsibility of a dedicated assistant, which passes the correct tool to the surgeon upon request and ensures that at the end of the surgery no tool or bandage is missing.
Another cart usually carries **monitoring and anesthesia** devices, which control the flow of drugs administered and keep the vital signs of the patient under surveillance. An anesthesiologist is commonly assigned to the task of intraoperative life support and pain management.
The walls of older ORs were often furnished with *illuminators* to hold photographic film containing medical images, such as radiographies. With the advent of digitalization, they have often been completely replaced by **screens** hanging on walls or mounted on the roof.

Fig. 1.3. **Open and laparoscopic surgery.** On the left, a modern operating room; in contrast to Fig. 1.2, a lot of attention is posed to the sterility of the environment. On the right, a laparoscopic setup; the additional ergonomic challenges on the surgeon can increase the difficulty of the most complicated procedures.
Image on the left under Public Domain. Elective surgeries hone surgical skills, prepare medical team for combat [Image 4 of 4], by Gloria Montgomery. The appearance of U.S. Department of Defense (DoD) visual information does not imply or constitute DoD endorsement.
Image on the right under Public Domain. Physicians perform laparoscopic stomach surgery, by Samuel Bendet over WikiMedia Commons.

Depending on the procedure, the most convenient location for the skin incision is chosen. Factors impacting this decision include considerations about the collateral damage and easiness of access to the target anatomy. The patient is commonly covered in a sterile cloth leaving a window sufficient to enclose the incision, which is performed after disinfection of the area. The seams of the incision and underlying muscles are pulled apart to provide access to the operating site. The tissues underneath may have to be resected in order to reach the actual objective of the surgery. This may consist of the resection of anomalous or cancerous tissue, or reshaping of problematic anatomy; alternatively, it may also be the delivery of therapy in the form of drugs, or placing implant. Finally, the original state of the surrounding tissues is restored to the maximum possible extent, by cauterizing or suturing. The last phase of the surgery usually includes the suture of the patient's skin.

1.2.2 Laparoscopic Surgery

As we mentioned in the previous Section 1.2.1, conventional open surgery requires large skin incisions to provide room and visibility for the surgeon to operate. The drawbacks in short- and long-term morbidity raised interest in how to reduce the invasivity of surgical operations. The invention and evolution of endoscopic techniques allowed to inspect the internals of the patient's body through just a small incision; concurrently, laparoscopic surgical instruments were developed in order to enable *minimally-invasive* manipulation of tissues. As a result, the physician can operate on the patient's internal anatomy by creating a few small incisions on the skin, and inserting through them the endoscope and elongated tools. In this setup, the damage and trauma to the patient's body is much more limited. While the definition **laparoscopic surgery** is perhaps the most known to the general public, this only refers to minimally-invasive surgery in the abdominal area; the same principles and, in large part, the same techniques can be applied to other regions of the body and are given a respective name (e.g. arthroscopy for surgery within joint cavities [120] or VATS for the thorax [94]).

As can be noted in Fig. 1.3, the setup of the operating theater remains largely unchanged with respect to open surgery, which also facilitates the reuse of the OR for multiple purposes. This can also be the case of a laparoscopic intervention, in the case of extreme complications which can no longer be handled with a minimally-invasive approach: the procedure is then *converted* on-the-fly to an open surgery to take care of the emergency.
In contrast to open surgery, the laparoscopic approach requires multiple smaller skin incisions. One gives passage to the **endoscope**, which is an elongated pipe serving as a hull to multiple fiber optic cables; one cable is used to deliver light to the operating site, while the other (or others in case of a stereo endoscope) serves to capture light in the visible range, which will be recorded by an RGB camera and then displayed on a screen for the surgeon. Having the light source and the cameras at the far end of the optic fibers helps with the miniaturization of the equipment and avoids thermal damage to the patient's tissues.
Further incisions are performed with the purpose of inserting the laparoscopic surgical instruments. Such tools are extremely miniaturized and typically feature a long, thin pipe between the handle and the actuator, so that they can be used to operate at a location deep within the patient's anatomy while holding them externally.
The surgeon can hold two surgical instruments with both hands, or one instrument and the endoscope. Control can be switched across phases of the procedure with an assistant. The assistant can hold the endoscope and move it according to the surgeon's needs, or assist from an additional port during particular moments of the surgery, for example when resected tissue must be removed from within the body. It is therefore easy to imagine that team coordination and communication are essential for the success, efficiency and error prevention during a laparoscopic intervention.

1.2.3 Robotic Surgery

The introduction of a large spectrum of robotic devices developed *ad-hoc* for a great number of procedures has undoubtedly contributed to improve patient care. However, the introduction of **robotic laparoscopic platforms** had a particular impact, changing the paradigm of surgery once again [107].

Such systems introduce one further level of **indirection** between the surgeon and the patient. In open surgery, the physician is in direct contact with the target anatomy and relies mostly on their own senses (vision and touch) to inspect it. Access to the surgical site and tissue manipulation are direct and intuitive.
Laparoscopy maintains the barrier between the internals of the patient and the surrounding environment intact; this means that the surgeon has to rely on a proxy of their own senses. Vision is redirected through a camera-monitor system, with the latter hanging in the proximity of the operating table, while interaction happens through elongated instruments with scissor-like handles. As a consequence, the surgeon must often not look directly at their own hands and the operating area, but rather at a screen positioned somewhere else. This fact, conjugated with the counter-intuitiveness of the action required to achieve the desired movement at the far end of the laparoscopic tools, poses a great cognitive burden on the surgeon and makes laparoscopy hard to master. The narrow field of view provided by the endoscope aggravates the situation even more.

Fig. 1.4. **Robotic surgery and microsurgery.** Robotic surgery and microsurgery share a distinctive feature: the surgeon's field of view is highly constrained by the visor employed to provide a visualization of the operating site. As a result, special care must be taken when integrating new devices into the workflow, in particular with respect to the user interface. However, robotic surgery improves the surgeon's hand-eye coordination over Laparoscopic surgery at the cost of increasing the distance to the patient.
Image on the left under Creative Commons 4.0 License (CC-BY-SA-4.0), by Fatemeh Dorosti over WikiMedia Commons.
Image on the right under CC0 License, from skeeze over Pixabay.

Robotic laparoscopy is an approach that promises to mitigate the drawbacks of laparoscopic procedures, while maintaining its benefits [23, 28]. As shown in Fig. 1.4, the distance between patient and surgeon increases further, even spatially: such platforms are shaped in a master-slave architecture, where the physician sits at the master console in order to control the movement of the slave robot. The master console is equipped with a visualization and an input system. The slave side features multiple arms, one holding an endoscope while the other arm is equipped with a laparoscopic tool.
The improvements with respect to conventional laparoscopic procedures lie mostly in the user interface presented to the surgeon. Video data recorded by the stereo endoscope is presented with a 3D visualization system, enhancing depth perception. The input system adopted by the most widespread platforms is pinch based, and allows a very intuitive control of the surgical instruments. The whole console is designed to give the native hand-eye coordination back to the surgeon, providing with the illusion of a point of view within the patient's body. Moreover, the operator can switch from controlling the arms holding the surgical instruments to moving the endoscope at any time, solving another of the major *pain points* of laparoscopic surgery. Some platforms offer not two but three arms with tool mounts, so that one can be used for auxiliary tasks like keeping one layer of tissues in a fixed position [43].

This approach was introduced recently, and continues to evolve at a fast pace. While the benefits for the surgeon are evident, the consensus about the improvements of patient care is not yet unanimous. However, it is easy to imagine the possibilities unlocked by the introduction of such a platform into the OR. We will discuss the topic in Chapter 4.

1.2.4 Microsurgery

The use of a microscope for the magnification of the operating area allows to operate on minute structures of the body, like small blood vessels and nerves. The technique was developed in the context of otolaryngology, but it can be naturally extended to virtually all the regions

of the body. Because of the fine manipulations it allows, it is particularly advantageous for reconstructive surgery or tissue transplant.

The setup, shown in Fig. 1.4, is similar to open surgery as presented in Section 1.2.1, but the surgeon's vision is mediated by a magnification device, typically an optical microscope. The magnification heavily restricts the physician's field of view, so continuous adjustment of the microscope position or zoom level is necessary. Typically, this is controlled directly by the surgeon through foot pedals in order to keep the hands free to operate. Such a setup is difficult to master, as it requires great coordination and independence of the limbs while maintaining a steady hand to avoid unintentional damage. The high level of magnification and the indirect view also make hand-eye coordination difficult. Adjusting to the former requires learning how to perform hand movements on an unusual scale, while the latter requires a mental mapping between the direction of the hand movement and the resulting effect visible through the microscope.

1.2.5 Image-guided Needle Biopsy

A biopsy intervention consists in the removal of a sample of tissue from the patient's body for examination. This is usually performed after an anomaly is located through non-invasive imaging with the purpose of further analysis.
A large spectrum of established procedures have been developed in order to cope with the many possible scenarios. Depending on the type, location and visibility of the lesion, different means can be used in order to locate, reach and sample it. A superficial skin anomaly may be entirely removed with a direct **punch biopsy**. Typically such lesions are visible to the naked eye, and the surgeon can simply remove a patch of skin with the aid of a dedicated sharp round instrument. The specimen can then be sent to microscopic or histological examination to determine malignancy.
If the region of interest is not located on the patient's skin, a more invasive procedure is necessary in order to reach it. In the most difficult cases, a **surgical biopsy** may be necessary to reach a very deeply nested lesion in a sensible area, which takes one of the forms described in the sections above. However, often a full-fledged surgery can be spared by using a *surgical needle* instead to reach the anomaly and resect a tissue sample.

A **needle biopsy** can usually be performed under local anesthesia with minor complications. However, many factors can have an impact on the final form taken by the procedure.
Depending on the type of lesion, different kinds of needles can be used: a *fine needle* (similar to a syringe) can be employed to remove liquid from an object such as a liquid-filled cyst, while a thicker *core needle* can be used to excise a specimen of solid tissue or bone marrow.
If the object to be biopsied can be felt through tactile inspection, as in the case of subcutaneous lumps, the physician can drive the needle by just relying on their senses; such a procedure is often described as a **freehand needle biopsy**. Otherwise, a mean of navigation is needed in order to successfully drive the needle into the target. Usually an imaging device is used to provide visual feedback, and the procedure can be hence categorized as an **image-guided needle biopsy**. Ultrasound is a natural choice in most cases for its interactive usage (as depicted in Fig. 1.5), but X-Ray Computed Tomography may be necessary in some challenging scenarios, such as the thoracic cavity [85].

Fig. 1.5. **Ultrasound-guided Lymph Node Needle Biopsy in the Axilla.** In this picture, the surgeon uses a hollow needle to remove a tissue sample from a Lymph Node visible in the ultrasound image for later histological examination. The physician must hold the US probe with one hand while introducing the biopsy needle with the other, and correlate the spatial information provided by the US image with the orientation forced by the particular setup while optimizing the probe position for visibility of both needle and target. This hand-eye coordination is not trivial and steepens the learning curve for US-guided procedures.
Image courtesy of SurgicEye GmbH and Dr. med. Dipl. med. Stefan Paepke.

While the reduced invasiveness of needle biopsy makes it an attractive alternative to more aggressive procedures (typically open surgery), the approach can pose a higher level of difficulty. The surgeon must handle one imaging device (or more, in case of multimodal imaging guidance) to image the region of interest and obtain a suitable picture of the operating site. In the case of ultrasound, the probe must be held with one hand. Concurrently, the needle must be inserted with the free hand. Both the surgical target and the needle must be kept visible in the image; this can be a challenging task, due to the low visibility of most surgical needles in the US image and the field of view of the imaging device, limited to a single plane. The need for coordination between the hands, the low understandability of the typical US image and the need to build a mental map of the surrounding anatomy contribute to raise the difficulty of the procedure.

Ultrasound-guided core needle biopsy of sentinel lymph node for breast cancer staging will be an application of the method proposed in Chapter 5. However, it is useful to consider such scenarios when evaluating the advantages and disadvantages of medical imaging and Computer Assisted Interventions presented in the next two chapters (2 and 3).

Part II

Introduction to Application Areas

2 Medical Imaging

The ability to obtain in-vivo structural as well as functional imaging of the human body is of crucial importance for the quality of patient care - from diagnosis to interventions to post-treatment monitoring. The availability of non-invasive imaging techniques, which can acquire such images from the outside of the patient's body, is a major enabler for new approaches which can significantly increase the precision and reduce the trauma related to interventions such as surgery and drug delivery.

This section presents an overview of the vast and advancing field of medical imaging. We will omit details about the individual modalities, which are easy to find in literature, in an attempt of providing a transversal picture of the topic which is less common to encounter. Instead, our focus will lie on how high-level characteristics of each imaging modality are reflected in their clinical adoption and usage.

2.1 Classification of Imaging Modalities

An imaging device can be categorized according to some orthogonal criteria, which we will go through in this section. While some of these properties are intrinsic to each imaging modality, others can be adapted in multiple variants to satisfy the clinical requirements of different scenarios.

2.1.1 Dimensionality: 1D, 2D, 3D, 4D Imaging

The output of the great majority of imaging techniques is a scalar or vector field: a grid of pixels or voxels, to which a value (scalar or vector) is assigned to represent the intensity of the measured physical phenomenon. This field is sampled (directly, or indirectly computed) over a corresponding region of space within the patient's body. The process can consist of a single acquisition, or it can be repeated over time for progressive refinement, interactive usage or for the monitoring of a physiological or surgical process over time.

The simplest case is **1D imaging**, in which the output is a single scalar value. An example can be found in ultrasound A-mode, which measures the echo generated along a single line, or in gamma probes, which can detect radioactive events originating from a region of space in front of them. Or also, by extension, in an electrocardiogram (ECG). Such a simple output can be rendered with an auditory feedback or the display of a number or plot, and is typically adopted in interactive or monitoring scenarios.

In contrast, **2D images** can be handled most intuitively, since they are closer to our own perception of the world and we are used to make use of them thanks to consumer multimedia technologies (photographs, etc.). They can be easily rendered on a screen, in particular in the case of scalar quantities: each pixel can assume any value within the output color space (e.g. [0,1] or [0,255] from black to white) via quantization and optional rescaling. However, in order to simplify the understanding of the image, the function between the scalar field and the final pixel values can be other than a simple linear mapping; this is hence often called an *intensity map*. A *color map* is an extension of this concept, as it arbitrarily defines a correspondence between the original scalar or vector field value range and the target color space (e.g. RGB), resulting in a colored visualization of the dataset.

Along with 2D imaging modalities, **3D imaging** is widely used for diagnostic purposes; less so for intraoperative applications, usually due to long acquisition times and risk of motion artifacts. These are due to the fact that 3D images are commonly acquired by progressive scanning of the region of interest, rather than simultaneous acquisition of all the data pertaining to a single image. Thus, motion of the patient or displacement of tissues during the acquisition may lead to an image which is inconsistent across its own extent. Furthermore, the domain in which the raw input is acquired is often different than the Cartesian space; the final scalar or vector field is then obtained through a reconstruction process. In order to be comprehensible for the physician, the most common visualization technique is the arbitrary reslicing of the volume into 2D slices. This operation requires the surgeon's interaction to select the desired view. Since their hands may be occupied or have to stay sterile throughout the procedure, the operator may have to ask an assistant to update the visualization for them. Such a workflow is not feasible for highly dynamic situations. On the upside, 2D reslicing of a 3D image can be however be done offline without further acquisitions, in contrast to 2D imaging modalities.

The repeated acquisition over time of 3D images is commonly referred to as **4D imaging**. The most widely known application of this technique is 4D echography, which can render spectacular videos of a fetus during pregnancy, but most other modalities have or are acquiring a 4D variant as well. Unfortunately some factors may limit the application of these techniques; for example, the high dose of dangerous ionizing radiation involved in a 4D X-ray Computed Tomography scan. Furthermore, one of the biggest problematics in their use consists in *motion artifacts* due to respiratory motion or heart beating during the acquisition of a volume; this issue can be mitigated by synchronizing the acquisition with these processes. More details about 4D imaging techniques can be found in an extensive review by Li *et al.* [101].

2.1.2 Energy source

In order to maintain non-invasivity, imaging modalities must be based on a physical phenomenon which can be detected from outside the patient body: these constraints are satisfied by waves (electromagnetic or mechanical) and subatomic particles.

Waves are typically transmitted to the patient by an **external** energy source. Among mechanical waves, those of ultrasonic frequencies are preferred due to their penetration capabilities and higher resolution. They can be excited mechanically by a piezoelectric array, or by light

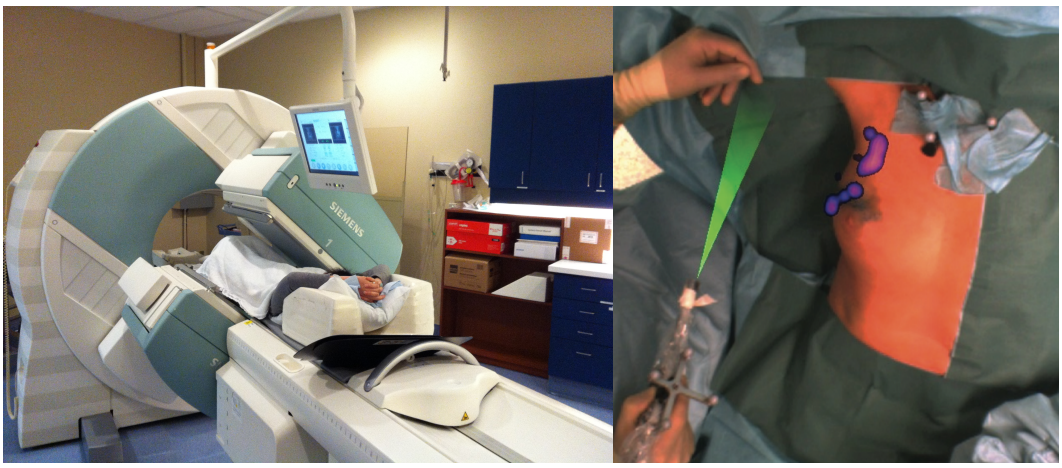


Fig. 2.1. **Gantry-based SPECT/CT scanner and freehand SPECT.** In this picture the contrast between pre- and intraoperative imaging is particularly noticeable. The left panel shows a diagnostic SPECT/CT scanner, which is designed with the aim of providing the best possible imaging quality. The shape and size of the gantry fit the patient's body so tightly that it would be very hard for a surgeon to interleave their workflow and acquisition of intraoperative images. On the other hand, freehand modalities allow such approaches; on the right, the output of an Augmented Reality system for intraoperative SPECT acquisition and visualization shows the field of view of a tracked gamma probe and the reconstructed distribution of the radiotracer within the patient's body.
Left image under Creative Commons License (CC-BY-SA 3.0), from ytrottier over WikiMedia Commons. Right image from [136], © 2012 IEEE

through the optoacoustic principle. Electromagnetic waves can be projected through the body, e.g. from an X-ray source, or excited by creating a strong magnetic field resulting in nuclear magnetic resonance.

On the other hand, particles must typically be delivered to the region of interest via injection or oral administration of a radioactive substance. Once this collects in the regions of the body which are the target of the examination, it constitutes an **internal** energy source which can be detected from a sensor external to the body, without further intake of energy. This approach is most popular among functional imaging techniques (and will be covered in Section 2.1.4).

2.1.3 Acquisition: Gantry vs Freehand

Most 2D techniques allow to sample the entire image simultaneously, being based on orthogonal or pinhole projection on a 2D sensor. An exception is ultrasound imaging, which is based on reflection. However, such a direct approach is not at all possible for 3D imaging, since a 3D sensor would have to occupy the same space as the patient. Hence, tridimensional imaging modalities employ tomographic reconstruction algorithms or progressive scanning of the volume of interest.

Tomography consists of the indirect computation of a scalar field through accumulated observations, such as measurements corresponding to its projection and integration along an axis. This process can be described as a linear equation system, and its solution constitutes an

inverse problem. In order for the problem to be solvable, it is necessary that the acquisitions satisfy geometric requirements.

This can be ensured via a fixed or moving **gantry**, holding the sensor and possibly the energy source. This approach allows for fast and accurate acquisitions, but often requires the use of bulky hardware.

The **freehand** method involves the usage of sensors that are hand-held by a physician and tracked in space by a dedicated hardware system. The operator has the responsibility to obtain a full coverage of the region of interest, possibly under the guidance of the system's software. Because of the dependency on the user, this approach is less widely used, mostly in an intraoperative setting.

2.1.4 Anatomical and Functional Imaging

An **anatomical imaging** modality can render an image in which the intensity of the pixels or voxels are correlated to a physical property of the anatomy within the region of interest. According to the physical principle on which the modalities are based, they can hence differentiate the tissues in very diverse ways and offer various representations of the same part of the body. X-ray based imaging modalities are particularly sensitive to variations in the density of the tissues irradiated during the process. Therefore, they are particularly apt to inspect bone shape and integrity. However, the resolution of modern equipment is such to allow the resolution of subtle variations in soft tissues as well. This fact does not make Magnetic Resonance Imaging, another diagnostic anatomical modality of preference for soft tissue, obsolete. The content of an MRI image makes other changes in the structure of tissues evident in contrast to CT, as it is more sensitive to the distribution of water than to density in itself. So, the choice of the imaging modality has a crucial impact on what will be visible to the physician.

Functional imaging does not attempt to provide an image of the patient's anatomy: instead, the resulting images provide spatial information regarding the localization of a physiological process of particular interest within the patient's body. Often this requires to inject a contrast agent into the patient's blood stream, consisting of a compound of a detectable substance (e.g. a molecule containing a radioactive isotope) and a substance which is involved in the physiological process of interest. As a result, the detectable marker will collect in the region of space where the process takes place. A functional imaging device can then provide a 2D or 3D image of the distribution of the marked compound in space, and hence of the intensity of the physiological process. A typical application is the diagnosis of several types of cancer, achieved through markers attached to sugar molecules. In these cases, cancer cells have a faster metabolic activity than healthy cells, so they collect and process more sugar which highlights them against the healthy tissue in a functional image.

A drawback of pure functional imaging modalities is that they only provide a scalar field rendering the distribution of the activity of interest in space, without anatomical landmarks useful to position the image seen into the patient's body. Achieving such alignment is then left to the physician, who has to correlate the shape of the colored dots on a black background to a

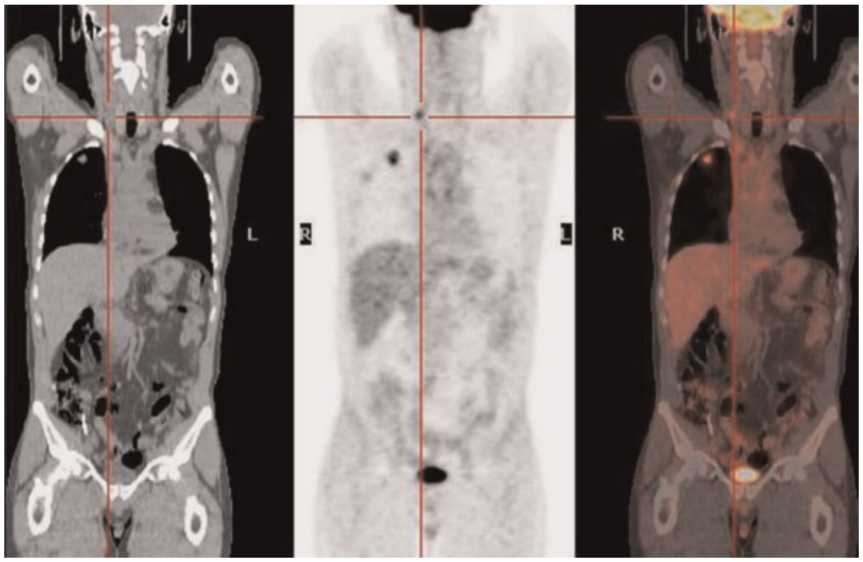


Fig. 2.2. **Example of multimodal image fusion.** This picture shows the importance of complementing functional and anatomical imaging for the comprehension of the whole image. While CT (left) is not capable of providing the functional information captured by PET (center), their fusion (PET/CT, right) allows to precisely correlate the position of functional information to fine anatomical landmarks.

CT or MRI image. **Multimodal Imaging** offers a way to avoid this difficult, time-intensive and error-prone process. It is based on the fusion of an anatomical and a functional image, resulting in an image where the intensity map of the physiological process of interest is overlayed on a complete rendering of the patient's anatomy, greatly facilitating its understanding. While it is possible to register a pure functional image to a previously acquired anatomical image, this can be a very difficult process with a high probability of partial or total failure. The convenience of acquiring the two images within the smallest possible time frame is so big (as shown by Fig. 2.2), that almost all CT scanners commercially available today integrate a SPECT detector. We discuss this topic more in Section 2.3.

2.1.5 Diagnostic and Intraoperative Imaging

Imaging modalities based on the same physical principle and acquisition, reconstruction and visualization techniques can differ greatly in order to satisfy different constraints. This can be recognized in the case of diagnostic and intraoperative modalities.

In a **diagnostic** setting, the highest priority is acquiring and visualizing the most accurate image, in order for the physician to take a decision based on the best possible information. For this purpose, the highest resolution and dynamic range are often pursued, while reducing noise and artifacts to the minimum. The least possible amount of information is also discarded (e.g. by restricting the device field of view), in order to provide as much context as possible. Considerations about the dimensions of the hardware, the time required for the setup and the acquisition, or other collateral factors have a lesser impact on the design of diagnostic imaging methods.

Priorities shift when entering the operating theater. **Intraoperative** imaging takes place at a time when the most important decisions have been taken and the plans laid out; its only purpose is to assist the surgeon in performing the intervention in the best possible way. To this end, the detail of the information may not be as important as its relevance for the task at hand, or the intrusiveness of the acquisition procedure within the surgical workflow. As a consequence, a real-time, easy to understand, high-contrast image can turn out to be the most helpful. An example is shown in Fig. 2.2; this topic will be further discussed later in Section 3.4.

2.2 Modalities

This section glimpses over the panorama of the most common image modalities, focusing on the implications that each technology has on their usage in a clinical setting. Our focus lies mainly on ultrasound and gamma imaging, which was employed for the contributions presented in Part III. A brief introduction to the other most common techniques is however also provided with the aim of providing a broader context to the reader.

2.2.1 X-ray based modalities

The first image of internal human anatomy acquired in a completely non-invasive fashion was a prompt application of the new radiation discovered by Roentgen in 1895, and by him denominated "X-rays". While he noted immediately that this kind of rays could travel through soft tissue but was strongly absorbed by bone, it was only one year later that this property was exploited in a clinical setting to inspect a fractured hand [185].

The methodology adopted to obtain the first X-ray images involved a radiation source and a fluorescent screen, between which the object to be inspected is placed. The X-rays (i.e. photons with wavelength between 0.01 and 10 nanometers) travel on a straight line between the source and the plate, unless they interact with matter on their path. In this case, the particles may be absorbed or scattered. The density of matter on the trajectory of the photons traveling from the source to each point on the surface of the screen determines the intensity of the interaction with the fluorescent agent contained in it. As a result, the produced images consist of the projection of the interposed tridimensional anatomy on a 2D plane, where high-density structures such as bones are rendered in a bright color against a dark background. This approach is hence called **projectional radiography**. While these characteristics make radiography particularly apt for orthopedic purposes, current technology allows for a sensitivity such that some sorts of variations and anomalies in soft tissues can be resolved as well, for example in lungs.
Radiography is also widely used in intraoperative settings for navigation and guidance, even if the dangers posed by intensive use of ionizing radiation such as X-rays requires proper evaluation of the advantages and disadvantages of its usage. In order to make intraoperative usage easier, the X-ray source and detector can be mounted on a *C-arm*, i.e. a semicircular support attached to a portable wagon or a rail on the ceiling, such that the patient can be scanned on the operating bed with a minor disruption to the surgical workflow. Many variants of 2D X-ray radiography find common adoption in current medical

procedures. A popular acquisition methodology, called **fluoroscopy**, consists of a continuous generation of images, which can be perceived as a real-time video. Low energy X-rays are employed in **mammography**, a widely employed screening and diagnostic method for breast cancer [185].

While the information provided by projective radiography can be invaluable, the provided images can be difficult to understand if the region of interest is cluttered: multiple layers of structures can be projected on top of each other, resulting in an equally cluttered image. This factor, together with the opportunity for better visualization, sparked interest into obtaining a complete 3D volume of the region of interest, rather than a simple 2D projection. This is achieved through a tomographic reconstruction, which was already described in Section 2.1.1. This modality is hence known as **X-ray Computed Tomography**, or also just as computed tomography or CT [76]. The advances in sensor technology and reconstruction algorithms make CT an extremely valuable instrument, capable of rendering internal structures in detail; however, the high dosage of radiation required limits its adoption.
The tomographic reconstruction process requires the acquisition of projective images from as many directions as possible; diagnostic CT devices achieve that with an assembly of source and detector which rotates at very high speed around the patient. For safety reasons, the hardware is enclosed by a bulky cover taking the shape of a tube, into which the patient is introduced via an electric bed (as shown in Fig. 2.1). This setup being unfeasible for intraoperative usage, a variant was developed with the common definition of **cone beam CT** (or CBCT). Typical CBCT systems involve a C-arm capable of autonomously rotating around their axis of symmetry for the tomographic acquisition to take place.

Both projective radiography and computed tomography can be augmented employing **contrast agents** [208], i.e. substances which significantly alter the interaction of an anatomical structure of interest with the physical principle the imaging method is based on, in this case X-rays. According to this principle, radiocontrast agents significantly absorb or scatter high energy photons and hence provide high contrast in radiographic images.
This method is for example used in **angiography** [61] to visualize blood vessels, by subtracting the images acquired after and before the injection of the contrast agent into the blood stream so that the structures present in both images do not occlude the objects of interest. Alternatively to injection, contrast agents can also be administered orally or in the form of enema, for the imaging of the gastrointestinal tract.

The biggest impediment to a wider usage of X-ray based imaging is arguably the dangers posed by **ionizing radiation** not only to the patient, but also to medical personnel. This is particularly important during a surgery, where the surgeon and the whole team must be forcedly close to the patient rather than behind a shielding wall as during the acquisition of a diagnostic image [174]. Since the risk of harm accumulates over the time of exposure to radiation, a surgical team commonly performing multiple procedures each day must adhere to strict regulations in order to minimize collateral irradiation. This involves keeping a safety distance from the device during image acquisition, and wearing very heavy shielding to protect the most sensitive parts of the body. However, the statistics showing a higher incidence of cancer in surgeons raise the doubt that these precautions may be insufficient [207].
Alongside safety considerations, a further drawback of X-ray imaging is the high sensitivity to metal objects which can lead to artifacts in the images, in particular after tomographic

reconstruction. Recent research in reconstruction and filtering algorithms helps minimizing the annoyance.

Despite these disadvantages, radiography remains a fundamental component of many diagnostic and surgical procedures thanks to the high resolution and contrast provided, as well as the speed of acquisition. The field continues advancing by refining existing techniques and investigating the adoption of further physical properties of X-rays.

2.2.2 Magnetic Resonance Imaging

Magnetic Resonance Imaging, or MRI, is an imaging technique capable of acquiring 3D anatomical volumes with high detail. In contrast to X-ray Computed Tomography, it is based on a physical principle which does not imply exposure to ionizing radiation. However, its commercial availability and rapid clinical adoption has reduced only partially the usage of CT. Reasons for this can be found not only in the high cost and (in the general case) slightly lower spatial resolution, but also in the different information contained in the image. Instead of displaying the density of matter in a given region of space, contrast in MRI images is provided by variations in the spatial distribution of particular atoms; clinical applications usually target hydrogen atoms, strongly present in water molecules and in other organic compounds. This makes MRI a favorite choice for diagnostic imaging of soft tissues and for functional imaging when an appropriate contrast agent is employed [3].

The physical principle exploited by MRI is **Nuclear Magnetic Resonance**, so the imaging modality was originally called *Nuclear Magnetic Resonance Imaging*. The increasing awareness of risks related to radioactive exposure caused the renaming of the clinical application in order to avoid confusion in patients: though acting on an atomic level, NMR is (at least at the current state of knowledge) considered harmless for humans.

The external energy is not delivered to the body as a localized, very high-intensity electromagnetic (EM) field which excite the atoms of a specific element present in the region of interest at a nuclear level. When the field is switched off, the atoms relax to their natural state and emit a radio frequency signal. By spatially encoding the intensity and gradient of the EM field, it is possible to correlate the frequency of the radio signal detected with the distribution of the phenomenon in the spatial domain and to reconstruct an intelligible 3D image.
Nuclear Magnetic Resonance is a phenomenon with various nuances, which can be exploited to extract alternative representations of the same anatomy where respectively different sorts of tissues are more evident. The modulation of the EM field can be used to measure the relaxation rate of the atomic nuclei along the parallel direction to the magnetic field, or in the transverse direction; these variants are respectively denominated T1 and T2. Further parameters of the process can be tuned to spawn an entire spectrum of *MRI sequences* of different diagnostic relevance.

Both CT and MRI are used for anatomical imaging, but the latter is also capable of functional imaging without a contrast agent. An example is provided by Functional MRI, or *fMRI*, which has gained public attention as an instrument for understanding the inner workings of the

brain by measuring the change in blood flow [3]. MRI contrast agents still exist for ad-hoc procedures.

Even though it is based on a completely different physical foundation, an MRI scanner is similar in appearance to a diagnostic CT device. Both have a bulky cylindrical gantry, which makes it their integration in intraoperative procedures difficult. The long acquisition time makes MRI particularly sensitive to artifacts caused by patient motion. Some patients might have difficulties lying still for a long time in a narrow tube, surrounded by the loud noises due to the constant switching of the EM coils. Breathing and heart beat can also generate perturbations of the image.
MRI however poses even more challenges to **intraoperative** use: the high-intensity EM fields interact strongly with any magnetic object present in the immediate surroundings, including most metallic objects. This not only forbids patients with any metal implant to undergo a scan, but also requires the whole Operating Room to be metal-free. This huge obstacle is overcome only by a few research institutions which investigate intraoperative MRI.
On the other hand, information provided by Magnetic Resonance Imaging can be essential for surgical procedures, for example to avoid critical structures during neurosurgery. This is often achieved by registering preoperative data to intraoperative images, or through stereotaxis.

2.2.3 Ultrasound

Ultrasonographic imaging (US) is a very popular diagnostic and intraoperative technique, being harmless, fast and cheap. It is mostly known to the public in the form of the echographic exam to monitor pregnancy, but its qualities make it a valid and flexible support for many clinical applications.

The acquisition of an US image consists in measuring the echo of an ultrasonic wave transmitted into human tissues by direct contact. This is usually achieved through a piezoelectric crystal, which can first convert an electric input into the source wave, and then perform the inverse operation when the reflected wave reaches the probe again. The resulting electric signal is digitalized and processed to obtain the final image. Since the wave reflection is determined by the changes in density of the tissues, this approach allows to outline significant structures in the region of space through which the mechanical waves travel.
The piezoelectric material is usually organized in a straight line, to form a 1D array. The shape can change in order to fit a particular anatomy or procedure. The sound wave travels orthogonally to the piezoelectric array, and when reflected reaches back to the sensor; by measuring the *time of flight* it is hence possible to obtain a 2D image. Multiple 1D arrays can be juxtaposed to build a 2D array, capable of synchronously acquiring a 3D volume; this approach is however still not as popular as the former. A cheaper alternative is a probe containing a 1D array mounted on a motorized support, able to rotate about an axis parallel to the probe surface. This makes the probe able to acquire a 3D volume without displacing it. A less common possibility is to translate the transducer array, rather than rotate it. An example is the Automated Breast Volume Scanner (ACUSON S2000™ ABVS; Siemens Medical Solutions, Inc, Mountain View, CA), which can acquire a 3D ultrasound volume of the whole breast after manual positioning by an operator. Preliminary results [203] report a very high sensitivity but low specificity compared to mammography. However, the elevated diagnostic

agreement between readers shows how 3D echography can reduce the dependency on the operator which is commonly associated to conventional hand-held ultrasound [58].
A further possibility for acquiring a 3D volume with a 2D ultrasound probe is offered by **freehand 3D ultrasound** (or *Ultrasound Compounding*). This technique allows to combine the samples acquired over time and space by a hand-held probe, constantly tracked by a mechanical, optical or electromagnetic tracking system. The main advantage is the improved repeatability of the measurement of the size of anatomical structures of interest with respect to 2D procedures, where the volume is estimated via approximated formulas based on measurements on notable planes intersecting the structure [54]. However, the quality of the resulting volume depends heavily on the skill of the operator and on the quality of tracking [188]. We will go deeper into the topic in Chapter 6.

The most common form of echographic probe is hand-held. Many variants are commercially available, with shape and properties adapted to single procedures or target anatomy. However, ultrasound probes have also been integrated with success into catheters [206] or endoscopes [45]. Lately a new probe has been presented for usage during robotic laparoscopic surgery (X12C4 Robotic Drop-In Probe, BK Medical ApS, Mileparken, Denmark).
Besides the shape of the probe, the reconstruction methodology can also be altered for various purposes. A contrast agent can be used to make objects of interest more evident in the image [86]. Or rather than on the intensity of the reflected waves as in standard **B-mode**, the focus can be set on measurement of the **Doppler** effect between the frequencies of the emitted and detected wave. This allows for example to measure the relative speed of the blood stream. Another successful application of ultrasound imaging is for **elastography**, that is the measurement of relative stiffness of the visualized tissues in order to spot anomalous regions.

Since US imaging makes use of a mechanical wave, the **penetration** capacity is more limited than CT or MRI. Most systems cannot reach further than 15 or 20 cm into the tissues. Moreover, since the wave must travel twice through the tissue before being detected, the chances for further collateral interaction with matter is higher and can result in image artifacts. Another consequence is that objects beyond a strong reflective surface (e.g. bones) will be reached only by the fraction of the wave getting refracted, being in practice *shadowed* and barely visible if at all. These two factors, along with a number of other physical effects make US images particularly noisy and hard to interpret, and limit its use in regions such as the thorax or the head (due to the obstruction of the rib cage and the skull respectively). Even worse, the aspect of an object may vary substantially together with the orientation of the ultrasound probe and the angle of incidence of the incoming mechanical waves.

All these factors make ultrasound imaging a very hard technique to master. Surgeons need extensive **training** before successfully performing US-guided procedures such as needle biopsies. Diagnostic exams are usually carried out by highly specialized physicians, sometimes called *sonographers*. These professionals often perform exclusively this kind of procedures, to the point of developing pathologies in their upper limbs due to the strain of maintaining proper contact between the probe and the patient skin [118].
The unique advantages of US imaging do outweigh its drawbacks in many cases. In particular, the dynamic and interactive acquisition protocol fits well to intraoperative scenarios, and allows the physician to inspect the anatomy during motion (for example of a body articulation),

while most other modalities require the patient to stay still to avoid motion artifacts. The lack of ionizing radiation involved in the imaging process makes US ideal for routine scans and for subjects at risk (e.g. women in pregnancy).

We will go into more detail about the 2D US B-mode image formation process in Chapter 5.

2.2.4 Nuclear imaging

All aforementioned imaging modalities involve an energy source which is external to the patient's body: a high-energy photon source in the case of X-rays, an electromagnetic field for MRI and a piezoelectric crystal for the generation of ultrasonic waves. This chapter describes two approaches that only rely on a hardware sensor, since the energy to be detected is directly emitted by the target anatomy; or, to be more precise, by a substance injected prior to the examination. Since most substances satisfying such constraint are radioactive, this kind of imaging modalities mostly falls under the umbrella of *Nuclear Imaging*.

By definition, these imaging modalities can only detect a contrast agent; so they are mostly used for functional imaging. The most straightforward approach, called **scintigraphy**, is to acquire a 2D image of the distribution of the contrast agent with a static external detector, as in the case of projective radiography. However, since in this case the source of the particles is not a point but rather internal to the body, the image is not formed through a pinhole projection as in radiography. The process is instead modeled as a parallel projection, due to the fact that the sensor is a flat panel containing a scintillator and a collimator orthogonal to the surface.

Nevertheless, the biggest difference between a scintigraphy and a radiographic image is the visualized content. While the latter provides a rather complete black-and-white depiction of anatomical structures, the former can be much harder to understand. The image may consist just of a colored stain on a black background, corresponding to the region of space where a physiological process of interest takes place. The lack of anatomical landmarks providing spatial context may make the result difficult to read. As a result, Scintigraphy is mostly used with contrast agents that bind to an entire organ, in order to inspect its shape and make anomalies evident to the physician's eye.

As in the case of radiography, **tomographic reconstruction** can be applied to multiple scintigraphic acquisitions in order to obtain a tridimensional image with evident advantages in terms of image readability. Two methodologies find most widespread clinical applications: Single-Photon Emission Computer Tomography (SPECT) and Positron Emission Tomography (PET). Even if conceptually similar, the two modalities are based on two different physical principles that reflect in how the acquisition and reconstruction processes vary between them.

SPECT contrast agents emit *gamma rays*, which are the most energetic electromagnetic wave observed until today. They are in fact much higher in energy than X-rays, and as a result even more dangerous after exposure. These rays can be detected through a *scintillator*, that is a device consisting of a chamber with a material prone to interact with the particle when

traveling through it. Typically this interaction results in the emission of an electromagnetic wave of lower energy, even in the visible range, which can be electronically detected. This mechanism allows to count the *events*, i.e. the number of instances in which a gamma ray interacts with the scintillator.
In order to filter out scattered rays or particles of other origin, a *collimator* is commonly placed in front of the scintillator cells. Consisting of a metallic regular grid parallel to the scintillator, it reduces the probability that an off-axis particle reaches the detector.
Given a number of scintigraphic images acquired from a set of orientations around the object, tomographic reconstruction can be applied to obtain a 3D volume. This can be achieved by using a gantry as in the case of X-ray CT and MRI, but also with a *freehand* approach, since small and portable gamma probes and cameras are available [115]. While these devices are usually employed for intraoperative navigation, when combined with a tracking system they can be used for intraoperative SPECT acquisition [142]. An example of SPECT/CT scanner and of freehand SPECT are shown in Fig. 2.1.

On the other hand, **PET image acquisition** requires a much more complicated hardware setup due to the particular manifestation of the observed physical phenomenon. While a gamma particle can be detected by a single scintillator chamber, hence limiting its possible origin to a straight line orthogonal to the gamma camera screen, this is not possible for positrons. These particles travel a very short distance since the time of their emission (typically less than 1 mm, depending on the isotope) before annihilating with an electron and in turn emit a pair of gamma rays traveling in opposite directions.
As a consequence, the standard approach to detect the emission of positrons is to identify couple of gamma rays traveling in opposite directions and having originated "simultaneously" (more sofisticated approaches take *time of flight* into account). This is achieved by organizing scintillators in a ring, and connecting them with dedicated hardware activated by simultaneous signals into pairs of scintillators. Once such correspondence has been found, it is possible to localize the point of annihilation along the straight line connecting the scintillators that reported the events. If the time resolution of the system is high enough, it is possible to restrict the probable origin of the event on a segment of limited length within the line.
This series of correspondences constitutes the input to the tomographic reconstruction algorithm, which has to account for the uncertainty around the exact position of the point of annihilation, as well as the completely unknown displacement between this point and the actual place where the original positron was emitted. All this factors limit considerably the spatial resolution that can be achieved with PET.

This drawback in terms of image definition, along with the high cost of operation of a PET scanner (and the associated particle generator, since the isotopes required have a very short lifetime) and the risk associated with the exposure to such highly radioactive materials, are still outweighed by the clinical value of the information it can provide. The same argument applies to SPECT, and explains the wide adoption of both imaging modalities in modern clinics.

More details about the usage of radiotracers for sentinel lymph node identification will be provided in Chapter 5.

2.2.5 Ongoing research

Even though the modalities introduced in the former sections are commonly available and adopted, medical imaging is a very active field of research and the panorama might change in the near future.

We already discussed in Section 2.2.3 the foundation of echography, which relies on the emission and subsequent detection of the reflected ultrasonic waves transmitted mechanically. **Optoacoustic imaging** (also known as *Photoacoustic imaging*) [179] makes use of ultrasound detectors to perceive ultrasonic waves generated within the tissue itself by irrorating the region of interest with light in the visible range. The application of intermittent light leads to rhythmic dilatation of the tissues, resulting in mechanical compression and decompression forming a detectable ultrasonic wave.
It has been shown that the absorption of light at certain frequencies is correlated to respective physiological processes, making optoacoustic imaging feasible for functional imaging. However, its reliance on light in the visible range poses challenges related to its penetration capability and the high scattering in living tissue, which in turn limits the spatial resolution.

Another imaging technique which is drawing increasing interest for possible clinical applications is **Optical Coherence Tomography** (OCT) [46]. This methodology makes use of *coherent light* to acquire images at very high resolution of structures directly below the tissue surface. While irrorating the area of interest with directed laser, reflected light is detected and filtered through *interferometry* in order to discard photons which have been scattered within the tissue, and to organize the remaining events according to their *time of flight*.
This technique is particularly promising for the manipulation of nanostructures, as in the case of microsurgery [38].

2.3 Multimodal Imaging

Functional imaging provides essential support for many applications, from cancer diagnosis to intraoperative decision support or navigation. The ability to visualize information related not just to anatomical structures but also to physiological processes invisible to the naked eye is a phenomenal enabler for precision diagnosis and surgery.

Unfortunately, it can be very challenging to interpret pure functional imaging alone, as suggested by Fig. 2.2. A radiologist, as well as a surgeon, needs to correlate functional information to anatomical landmarks or structures in order to make an informed decision. Therefore, there is large interest in **multimodal imaging**, i.e. techniques capable of coherently visualizing image data acquired through different modalities.

2.3.1 Image Fusion Strategies

Most modalities previously discussed in Section 2.2 have reached wide availability in modern hospitals. Therefore, there is extensive research attempting to jointly visualize images acquired

through independent devices, for example a SPECT and a CT acquired at different points in time. A precondition to such task consists in aligning the two images, such that the information contained in respective voxels corresponds to the same region of space in the patient's body. This task is referred to as **image registration**, and is an extensive and active field of research [64, 110].

In general, it is not possible to guarantee that the patient assumed the same position during both examinations. The respiratory phase can also introduce extremely complex compression and deformation of the patient's soft tissues. As a consequence, the correspondence between voxels of the two images can often not be formulated as a **rigid** Euclidean transformation; or to be exact, as a function at all (since one voxel in one domain can correspond to more than one in the target domain). However, for rigid structures this assumption can be made to simplify the problem of **deformable registration** greatly. Otherwise, *a priori* knowledge about the problem can be used to build a mathematical model of the imaged anatomy and limit the solution space to a set of parametric deformations.

Once the solution space has been identified, for example $SE(3)$ in the case of rigid transformations, the problem is usually formulated as the minimization of an *error function*. A popular choice is to minimize an expression involving the numerical value of the voxels of the two images; this approach is hence often called **intensity-based registration**.

Another possibility is to identify anatomical **landmarks** in both images, in order to derive the global transformation from their relative displacement. In order to increase the precision obtainable with this method, some approaches involve placing artificial landmarks (often called **fiducials**) providing high contrast in both images; an example can be found in stereotaxy for neurosurgical navigation.

2.3.2 Native Multimodal Imaging

The registration of two images acquired independently remains a very challenging problem, as the possible deformations constitute a very large solution space and the difference in appearance of the same object in the two modalities can make it hard to find correspondences between the images.

Image registration can be made superfluous (or reduced to a known calibration transform) if the two modalities are **acquired in parallel** by a combined device. Usually this is achieved by placing the sensor hardware for the respecting modalities next to each other in a single gantry; in this way, the common workflow involving the patient being translated through the device via an electric bed is not disrupted. Since the relative position of the two sensor arrays does not change, it is then possible to calibrate the device once to merge the two imaging modalities into a single multimodal image. If the two sensors are close enough, even the artifacts due to breathing or involuntary motion of the patient during the examination can be minimized.

The most popular diagnostic multimodal imaging devices combine an anatomical and a functional imaging process within the same hardware. This allows to provide anatomical

context to the functional image, with the result of improving its readability and improve the accuracy of the diagnostic process. The first native multimodal device combined X-ray CT and SPECT; today this combination is so popular that it is very hard to find a commercially available pure CT scanner, since most producers offer only combined **SPECT/CT** devices. The same principle led to the development of **PET/CT** and **PET/MRI** scanners [26].

The possible insights unlocked in research and clinical applications by multimodal imaging keeps the subject an active field of research. It is currently being attempted to develop single sensors capable to act as detectors for both modalities, in order to further reduce motion artefacts. The feasibility and clinical advantages of combining different anatomical modalities, such as CT and MRI, is also being investigated. Furthermore, contrast agents detectable by multiple modalities are also being researched, in order to more accurately correlate information from both sources [26].

2.3.3 Applications of Multimodal Imaging

Native multimodal imaging offers better image quality and registration accuracy. Therefore, it is widely employed for diagnostic applications, where quality and accuracy are paramount for the physician to make accurate decision and, in case of subsequent intervention, planning.

However, as discussed in Section 2.1.5, priorities shift in an intraoperative scenario. As single modalities had to be adapted for usage in the OR, multimodal imaging can as well only be applied if the disturbance to the surgical team's activity is outweighed by its benefits. With this premise, much effort is being spent in order to find approaches capable of exploiting information from high-quality preoperative imaging during surgery, for example by registering the output of a flexible and fast modality such as ultrasound with a preoperative MRI image. We will talk more about intraoperative multimodal imaging in Sections 3.3.1 and 3.4.

2.3.4 Conclusion

In this chapter we discussed the usefulness of multimodal imaging, in diagnostic scenarios as well as for intraoperative support. The ability of differentiating tissues not only on the basis of their anatomical structure but also depending on ongoing physiological processes can be of terrific value, and this fact is driving both research and product development efforts.
We also discussed the constraints that imaging devices must satisfy in order to enter the OR in contrast to diagnostic setups, and how these constraints shaped up the design of intraoperative variants to significant deviations from the original concept.

These factors were crucial for the elaboration of the concept of our multimodal imaging system, that we will present in Chapter 5, and motivate our effort to improve the usability of freehand 3D ultrasound (which is the subject of Chapter 6).

3 Computer Assisted Interventions

The previous chapter discussed the most common medical imaging modalities. Such image data can be used for diagnostic purposes, unchanged or after further manipulation with Computer Assisted Diagnosis (CAD) methods. In this work we will focus instead on methods for assistance during surgery, often referred to as Computer Assisted Interventions (CAI). A consistent portion of said approaches make use of imaging data, and as such satisfy the definition of Image Guided Therapy (IGT).

3.1 Definition and Objectives of CAI

The application of electronic and computing technology for support of surgical tasks is a large and heterogeneous field, in its means as well as in its purposes. We will discuss the most widespread in this section.

A recurring topic in this context is the reduction of **invasivity**. The minimization of the number and size of incisions on the patient's skin is directly correlated to the outcome in terms of postoperative morbidity, recovery time and life quality. Damage can also be minimized in terms of collateral alterations to the the functionality of the nervous system or other vital organs, which can lead to lifelong consequences for the patient. These two aspects can reflect into health as well as economical benefits for the individual patient and the whole society [107]. The minimization of invasivity can be a consequence of the more general reduction of guesswork and error that can be achieved by employing hardware and software methods [81]. Advanced visualization, navigation and guidance, and collaborative robotics can help completing the surgery with success and confidence.
Furthermore, the adoption of new technology can enable **new procedures** that would not be possible at all otherwise. As the invention of endoscopy opened the door to today's laparoscopic surgery, a stream of innovation continues entering the whole field of surgery, making new approaches feasible and improving quality of care.

While the patient is of course the first beneficiary of most advances of modern medicine, there is another stakeholder which is often undeservedly forgotten: the surgeon. New technological applications can reduce the **danger** posed by surgical procedures to the surgical team. An example is the exposure to ionizing radiation during X-ray guided interventions. By tracking the position of the X-ray source and the personnel within the OR, it is possible to ensure that the entire team is outside of the area of direct irradiation [105]. Another possible area of improvement is **ergonomics**, in particular of laparoscopic surgery. Multiple studies [90, 147, 187] stress the hard conditions posed by laparoscopic setups to surgeons. The development of new visualization techniques and instruments, as well of robotic surgical systems, can help greatly to mitigate the problem. Along the possible insurgence of health problems caused by

ergonomics, current OR setups can induce **high stress and cognitive load** in the surgeon, which are found to be correlated with a decline in decision making ability and surgical accuracy [10, 24, 84].

3.2 Techniques of Computer Assistance

Here follows a brief overview of the extensive field of CAI. For the sake of brevity we will limit our scope to a conceptual map of existing approaches and techniques, since exhaustive reviews can take the shape of whole books [81].

3.2.1 Tools and Devices

Virtually every modern OR is equipped with multiple **displays**, showing various information about the patient and the status of the surgery. Preoperative data can be visualized on multiple screens surrounding the operative site, for reference to the surgeon. The output of the endoscope can be visualized for both the surgeon and the whole team; this allows the assistants as well as medical students present on site to follow the surgery. As technology progresses, currently used LCD screens may be replaced at least in part by 3D capable displays or head-mounted displays (HMDs) in order to provide spatial information and facilitate the surgeon's hand-eye coordination [136].
The **imaging devices** can also be counted among the technologies that are made available to the medical team. All imaging modalities have by now taken a digital form, and offer a wide range of customization possibilities in order to find the most optimal configuration for the operating surgeon's tastes and the nuances of the surgery at hand.
In order to exploit the available pre- and intraoperative imaging data for interactive assistance, it is necessary to be able to correlate the position in space of the surgical instrument and of the image volumes or planes with respect to the patient. This can be achieved by using **tracking systems**, which usually take the form of a stationary device (e.g. a camera) following markers attached to objects of interest (in our example, planar markers bearing grayscale patterns). After calibrating the tracking system together with the employed imaging devices and surgical instruments, it is possible to derive the position of an instrument within an acquired image for navigation support and other applications.
With the recent developments in the area of computer vision enabled by deep learning techniques, **markerless tracking** approaches have reached levels of accuracy and performance that may allow them to be employed soon within the OR. In contrast to marker-based tracking, such methods are able to recognize the object from its own appearance within a standard RGB image. If a high-definition video stream is available, it becomes therefore possible to extract the outline of the object and to estimate its pose in space with respect to the camera. This constraint, together with the *line-of-sight* requirement, makes these methods a natural fit for robotic laparoscopic surgeries, where the workflow already provides a close-up video feed of the operating site and occlusion is avoided by the operator for their own convenience [92].
With these techniques, it becomes possible to employ simple color cameras to acquire knowledge about the world. However, in recent times technologies capable of providing 3D information have become available even in the consumer market. RGB-D cameras, which associate to each pixel not just a color value in the RGB spectrum, but also a depth value expressing

the distance from the camera to the closest visible object along the associated ray, have found applications in a vast spectrum of research and industrial fields. Their miniaturization allows them to be integrated into headsets capable of mapping the environment and detect user gestures, such the HoloLens (Microsoft Corporation, Redmond, USA). In the medical field, their usage has been explored for intraoperative Augmented Reality [62, 136] and for integration within endoscopes [67, 164].
Rather than guiding the surgeon with visual or auditory feedback, once spatial relationships are known it is possible to employ **robots** for surgical assistance. The degree of autonomy can vary: the robot can perform a task completely autonomously or just steady the physician's hand, or guide it to a predefined point in space within the patient's anatomy. Robotic devices and applications will be discussed in Chapter 4.
As the number of devices within the OR increases, the task of managing them in the context of the ongoing surgery risks to increase in complexity as well. As a consequence, the optimization of their **user interface** acquires relevance with the aim of reducing the mental workload of the surgical team, and with it the probability of error and the operating time. We will discuss this topic in more detail in Section 3.5.

3.3 Applications

The tools presented in the previous section can be employed for surgical assistance in various ways. Here we will explore four directions in which CAI technologies can support the surgeon: "seeing", "guiding", "assisting" and "knowing".

3.3.1 Seeing: Imaging

CAI directly integrates many intraoperative variants of the imaging modalities presented in Chapter 2, and makes large use of preoperative diagnostic data.

In Section 2.3.1 we introduced the technique of **image registration**, which is useful to bring two different images into the same coordinate system and spatially align them. This can not only assist in the creation of a multimodal image from a functional and anatomical pair of images, but it can also be exploited to support interventional navigation and guidance. While intraoperative imaging alone is not always sufficient to this purpose because of the low quality and/or field of view, the data can be often used as a base to which the high-quality preoperative diagnostic data is registered to. After registration, the spatial correspondence between the position of the patient's anatomy and the preoperative image is known and can be used for intraoperative assistance, like navigation, guidance or interactive advanced visualization.
More advanced approaches require a semantic understanding of the content of the image. This can be at least in part achieved with **image segmentation** techniques. Segmentation consists of assigning a class or label to each of the image pixels/voxels. Examples of useful segmentations may find the boundary of a particular organ or tumor against the background in an anatomical image like CT or MRI; or delineate the regions of the brain in a volume of the head, or recognize vessels in an intraoperative ultrasound image. Once the spatial location of an object of interest is known, this enables augmenting surgical workflows with interactive

navigation (see the next section) or robotic guidance with active constraints (discussed in Section 4) to reach or avoid it.
Diagnostic and intraoperative images can be used by the surgeon for **planning** the intervention. Being able to build a mental map of the anatomy surrounding the operating site before taking action can prevent errors, reduce stress on the surgeon and minimize collateral damage. To this end, advanced visualization techniques are being investigated to present the most relevant information [168]. Another possibility is 3D printing the imaged anatomy, which provides a model that can be interactively explored by a surgeon before starting the surgery [102].

The usage of intraoperatively **registered** diagnostic data has been adopted with success for orthopedics and neurosurgery, as we will discuss further in Section 3.3.2. However, its wider application to general surgery is progressing more slowly. One of the reason for this lies in the complex deformations that soft tissues can undergo under compression or resection during surgical procedures or even simple patient movement [11, 173]. Since soft tissues are subject to free deformation rather than simple rigid displacement, correctly registering a preoperative image or volume to the current anatomical configuration has two preconditions which are still difficult to achieve at the present state of technology: information about the said anatomical configuration, and a reliable registration procedure. As stated in Section 2.3.1, the latter is still a hard and interesting problem. But it is aggravated even more by the constraints posed on imaging devices to fit into the operating theater. As we anticipated in Section 2.1.5 and we will discuss in more detail in Section 3.4, the large diagnostic devices can not be used to monitor the patient's anatomy in real time because of their dimension, configuration and acquisition time; their reported "intraoperative usage" typically consists in temporarily halting the surgery to acquire an image. This leaves us with intraoperative variants such as CBCT or ultrasound, which are only capable of 2D imaging or are severely limited in image quality. As a result, their usage as reference for deformable registration still is an active research topic [143, 177].

A breakthrough in this field would enable advanced usage in the OR of preoperative images and derived data (e.g. after segmentation of the image) which could not be obtained from intraoperative data alone. One example of such opportunities can be found in the application of *Augmented Reality*, which could be particularly useful during endoscopic procedures where the surgeon's field of view is limited and the cognitive burden posed for spatial orientation is extremely high. This factor accounts in large part for the long training needed to master laparoscopic and robotic surgery [6, 11].
In the meantime, the problem of disorientation in endoscopic procedures due to the narrow visual field is being tackled in simpler ways, such as **mosaicking** [170]. By matching local features between successive video frames captured by the endoscope, the displacement between them can be computed and a global panoramic map can be built. Such map can be useful as a support for the surgeon to orient him/herself in particularly challenging scenarios. It is however possible to use the same information to perform a **3D reconstruction** of the environment, rather than just a bidimensional image. This task can be achieved through the direct acquisition of 3D information with an RGB-D endoscope, or by exploiting local features of the image to perform Simultaneous Localization and Mapping (SLAM) in order to simultaneously derive the position of the camera within the environment, and the shape of the latter [108]. This task can be made easier through the adoption of a stereo endoscope, which is common for robotic laparoscopic platforms. Once a 3D map is built, it may be used

for registration to preoperative data (with the purposes outlined earlier in this section) and/or for implementing Augmented Reality. However, the high dynamicity of the environment due to the deformation of the soft tissues in the abdominal area and the surgeon's own activity makes such tasks difficult, and requires a continuous update of the 3D map.

3.3.2 Guiding: Navigation

In general, the objective of surgery is reaching a target region inside the patient's body to remove or reshape tissue, to place an implant or to deliver treatment locally, often in the form of a drug [81]. As a consequence, a substantial part of the time and effort of the surgical procedure is actually spent while opening a path to this target region for the surgical instruments, and then closing it back to minimize permanent damage. Optimizing this path can lead to a significant improvement of the patient's life quality in the postoperative phase, as well as reduce the operating time and resources required. Surgical planning based on image data, as described in the previous section, can help prepare the surgeon in advance. But interactive navigation constitutes a direct form of assistance during the intervention itself.

Conceptually, navigation consists of providing a real-time **feedback** about the spatial location of an instrument with respect to a target position or orientation. Usually this information is provided visually, by drawing the position of the surgical instrument within a visualization of a pre- or intraoperative image or also possibly through an abstract user interface involving geometric primitives.
The navigational interface is commonly rendered on a screen positioned close to the patient bed. It is also possible to employ augmented reality to reduce the distance between the operating and visualization sites and align them spatially, with the aim of facilitating hand-eye coordination and hence reducing the stress posed on the surgeon. Recent availability of advanced HMD technology rose interest in this approach again, but the weight of even the newest devices still makes them uncomfortable for prolonged use while looking down, which is the most common position assumed by the surgeon during procedures [11].
However, other means of communication are being investigated, such as auditory feedback as in *Sonification* [112]. The aim is to increase the quantity of information that can be concurrently provided to the surgeon, without forcing him/her to, for example, switch the focus of their vision between different regions of a display or between the operating site and a screen.

Meaningful navigation has numerous preconditions. A typical pipeline involves correct knowledge of a chain of geometrical transformations: from the patient to the image on which the position to reach or the region to avoid has been defined, to the tracking system being employed in the setup, to the marker attached to the surgical instrument, finally to its tip (or, depending on the instrument, its active region).
Patient-to-image registration can be achieved by using *fiducials*, which are objects that can act as identifiable landmarks within the image. Usually, this is achieved by choosing a material which is rendered in a particular region of the image's dynamic range, such that their segmentation within the image becomes trivial. After segmentation, their coordinates within the image's reference system can be easily computed. Generally, such fiducials also act as markers for a tracking system, so that the transformation between the patient and the tracking

system, and in turn between the image and the tracking system, can be derived. However, placing fiducials can be an invasive procedure; this justifies the research for *landmark-free* registration techniques, mostly implemented using intensity-based methods.
Since the transformation between the tracking system and the marker attached to the surgical instrument constitutes its very output, the transformation between the marker and the active region of the tool remains before closing the chain. Determining this quantity via a proper **calibration** process is vital to achieve high accuracy [72, 131]. A faulty computation of the translation between the marker and the instrument will introduce a systematic error between the expected and the real output. A perturbation in the rotational component of the calibration matrix can be even worse, as it will lead to larger and larger displacements between the output of the system and the real world as the distance from the tracking marker increases.

3.3.3 Assisting: Robotics

Imaging devices and tracking systems provide means to obtain a digital representation of the real world, which can then be virtually manipulated and presented to the user with computers. Robotics closes the circle, allowing to affect the real world based on this information. Medicine is a suitable target for the use of robots to automatize procedures or provide collaborative support to the operating surgeon. Here we will briefly present how such approaches fit in the CAI context, before discussing the topic in more detail in Chapter 4. Extensive reviews of the topic can be found abundantly in literature [36, 182, 184].

A robot can be generally described as a system capable of moving in space, by changing its position in the environment or modifying its own geometric configuration. Often the purpose is to move a tool attached to the robot, or more precisely its *end effector*. The robot movement can be directly controlled by a human using a suitable user interface, such as joysticks, or otherwise the control can be assumed by software routines, usually targeted to a well defined task.
The end effector of the robot can for example be a surgical instrument, such as a scalpel, a cauterizing device, or a biopsy needle; the purpose of this kind of applications can be exploiting the steadiness of the robot movement. Even the most skilled surgeon is not immune to jerkiness in the movement of the hand, so systems apt to its mitigation can be highly beneficial, in particular in case of operating sites cluttered with delicate and critical structures (e.g. vessels or nerves).
Alternatively, by coupling the robot system with an appropriate sensor, it is also possible to synchronize the movement of the robot with the patient's breathing or cardiac cycle so that the surgeon does not have to compensate for it. Commercially widespread robotic platforms for laparoscopic surgery can absolve both these tasks, as well as other specialized systems.
In other applications, a sensor or imaging device is attached to the robot, such as in the case of **endoscope holders** [166, 197]. In a standard laparoscopic setup as described in 1.2.2, the surgeon must either use a hand to hold endoscope in order to keep the operating area in the field of view of the camera, or let an assistant hold it. The latter option is of course very challenging for the assistant, which has to carefully follow the physician's commands and the procedure as visible on the screen in order to foresee the next movement. Each surgeon has an own "style" and preferences, and so developing good team coordination requires time and effort. As an alternative, automated robotic system have been developed and are currently

being further investigated to replace the assistant. The aim is to achieve better efficiency by using implicit input as the position of the surgical instruments, the surgeon's gaze direction or semantic knowledge about the current phase of the surgery, recognized through e.g. the video stream.

Robotic systems with a varying degree of **autonomy** have found both research and clinical usage. However, costs, benefits and risks must be carefully evaluated in order to find the optimum for each procedure. The development of a fully autonomous device involves the design of dedicated hardware and software, which must undergo a very rigorous certification process before it can be allowed to operate on a living patient under only passive supervision. Mechanisms for acquiring continuous feedback about the ongoing procedure and for detecting unforeseen problems must be integrated, to avoid irreparable damage. Furthermore, the legal and ethical implications of a fully autonomous approach must be weighed with care.
Lesser degrees of autonomy can be more reassuring for most medical practitioners, hence leading to higher and faster acceptance. Mechanisms under direct control of an operator do not need to provide a replacement for the human cognitive capabilities, which is today still a daunting task. The focus then shifts on complementing strengths and weaknesses of humans and machines, by using the latter as "extension" of the former, or by implementing a collaborative behavior between them.

We will resume the topic in Chapter 4.

3.3.4 Knowing: Surgical Data Science

In the last section we stressed the difficulty of creating a fully autonomous machine capable of human judgement, and the relative convenience of rather exploiting the precision and steadiness of mechanical movement to improve the surgeon's skills or, in other words, to give them "superpowers". A recent trend is arising which has a purpose similar in concept: augmenting human intelligence with the superhuman memory and processing power that can be provided by computers.

The difficulty in building a shared knowledge base and transmitting information is known even on a local scale, such as within the same institution [132]. As a consequence, the variability in skill and expertise among surgeons is substantial. This could be mitigated by building a knowledge base for real-time retrieval and proactive visualization during surgery. The information could be provided explicitly in surgeon reports, or derived in an automated fashion by examining data collected by sensors distributed into the OR.

Such a task requires **modeling and understanding** of surgical procedures. Surgical Process Modeling [93] has been a very active field of research in recent times in virtue of its possible applications for intraoperative assistance, decision support, clinical training and performance evaluation. Recent advancements in AI triggered by deep learning techniques represent a milestone in the progress of machine perception, in particular for computer vision where benchmarks report accuracy in detection and recognition surpassing human performance. However, current approaches still largely rely on local features of the image and lack contextual and semantic knowledge of its content, as can be shown through simple experiments [160].

While this still makes such techniques an insufficient foundation for unsupervised autonomy in critical applications, they still constitute an extremely powerful tool for information retrieval and decision support which can aid a human operator. By examining the laparoscopic video feed, the output of intraoperative imaging and life monitoring devices in real-time, and based on preoperative data such as diagnostic imaging, it should be feasible to retrieve information about similar cases or detected anomalies which can help the surgeon in making informed decisions or adapt the surgery to the current situation.

Many current efforts in Computer Assisted Diagnosis (CAD) attempt at assisting the physician during the diagnostic process by suggesting regions of interest based on computer vision techniques, but also by searching the available knowledge base for previous similar cases to compare with. The search criteria can be information derived by the image itself, as well as personal information about the patient which can affect the diagnosis (such as gender, age and medical history). Current laparoscopic setups already provide a live video stream of the surgery, which already on its own constitutes a very valuable input for such approaches. Their use could enable real-time, context-aware assistance during the intervention.

Recently, a formal definition of these approaches has been proposed by a consortium of leading researchers under the name of Surgical Data Science [109].

3.4 Properties of Intraoperative Imaging Modalities

Laying at the foundation of CAI and IGT, and constituting the focus of this work, intraoperative imaging deserves particular attention in this context. In Section 2.1.5, we already mentioned the different constraints posed on intraoperative imaging devices and methodologies with respect to preoperative diagnostic scenarios. This dichotomy is recognized by multiple authors [81, 137], and can be summarized in some recurring observations:

Specificity Diagnostic devices are mostly general-purpose oriented, as the problem of diagnosis is by definition open-ended. CT and MRI are generally implemented as full-body scanners, although providing protocols for imaging more restricted body regions. Intraoperative devices must be able of targeting the anatomy of interest and render the most *relevant* information for the procedure at hand.

Usability/Quality Tradeoff In the diagnostic phase, accuracy and completeness of information is paramount. The physician reading the image must be put in the best possible position to take an informed and accurate decision; considerations about the convenience of the **setup** and the acquisition **time** assume a secondary role. However, during surgery, priorities are typically inverted. In order to minimize morbidity and chances of error, imaging should be integrated into the surgical workflow in the least disruptive fashion. This involves reduced acquisition time and, when possible, real-time feedback. While it is crucial that the reported information is correct and complete, the definition of "completeness" changes: an intraoperative image is complete when it contains the minimal information viable to achieve the surgical task without error. To this end, a low-quality image can suffice if it provides a high contrast representation of the objects

of interest. The same principle does not apply just for the acquisition speed, but also to the **size** of the device and its shape: a large gantry where the patient fits barely inside cannot be used during most procedure without coming to a complete halt and a total disruption of the workflow. Collateral characteristics of the imaging device or process factor into this tradeoff. Sensibility to **magnetic objects**, as for MRI or EM tracking, requiring a metallic-free operating scenario complicates greatly the integration of such approaches.

Safety A reason for the low pace at which innovation reaches mainstream adoption into operating theaters can in large part be found in safety and reliability considerations. Total failure, such as a system shutdown, is not admissible if it puts the patient's health in danger. Partial failure, such as not rendering complete or correct information, can be even more dangerous as it may be subtle and difficult to recognize, and in this way jeopardize the clinical outcome of the surgery. The exposure to hazardous substances or emissions potentially harmful for the patient and the clinical team is another factor which can limit the cases in which an imaging modality can be made use of. The clinical benefit brought to the outcome of the intervention must outweigh the inconvenience and danger.

In light of these considerations, it is easy to understand the preference of C-arm radiography and CBCT (already introduced in Section 2.2) in intraoperative scenarios over full-fledged CT scanners, in spite of their superior performance. The same applies to SPECT, which can be found in some operating theaters in its freehand variant rather than as a dedicated gantry. Some research facilities host an MRI scanner in a metal-free OR, but this approach has not translated yet to mainstream adoption. Finally, the speed and flexibility of echography fits well for guidance support within a large class of interventions, to the point that *drop-in* variants have been developed for laparoscopic usage. Laparoscopic applications of other modalities, such as SPECT [52] or optoacoustics [125], are still in the research stage.

3.5 Acceptance of CAI

The integration of new technologies in surgery is naturally slow, due to the criticality of the mission. New devices and approaches can only be introduced with the utmost prudence in order not to endanger the patient with long-lasting damage or life loss. As a natural consequence, innovation takes mostly form of incremental updates, and profound changes to the overall workflow are rare. Such disruptions can only take place when the benefits overwhelm the possible risks and the learning fatigue. Furthermore, as more and more devices find their way into clinical routine, the risk of cluttering the OR increases. With the advancement of technology and of minimally-invasive surgery, the **complexity** of the processes and of the involved technology raises as well. In order to contrast this trend, a holistic approach is required to encompass the whole surgical process and avoid getting caught up in the own *silo* of competence, may it be medicine, engineering or an even narrower niche. These problematics have been brought to attention by some authors [144, 151].

Multiple studies conducted in the last years [24, 80, 87, 117, 202] correlate the **mental stress** of the surgeon, as well as the efficiency of communication within the team and disruptions of the workflow, with clinical performance. All of them state that further investigation is needed before making strong claims, in part because of the difficulty in proving causation rather than correlation when studying complex scenarios. Still, they report trends in this direction. Hence, it can be hypothesized that research aiming at strongly reducing the surgeon's cognitive burden may help them focus on the actual surgical objective. Therefore, streamlining the tools adopted to handle imaging data or to support the decision process could in turn significantly improve the clinical outcome.
Other authors [98, 147] warn about another danger that is often undervalued: the lack of **ergonomy** of minimally-invasive surgical procedures, in particular laparoscopy. In a setup such as the one described in Section 1.2.2 the surgeon is forced into an inconvenient position, which must be held for long time (even hours) in order to avoid harm to the patient and to avoid spending excessive time in the intervention. These studies report a big incidence of chronic pain in surgeons operating laparoscopically, in particular in the back and upper limbs. The little awareness among the physician themselves about the problem and the existing strategies to minimize it show once again the inertia existing in the field and the opportunities for improvement.

3.6 Research Directions

As the quantity and quality of information available during interventions continues to increase, new possibilities open up for novel technologies and applications capable of effectively supporting the surgeon in exploiting and managing said information. Laparoscopic surgeries would be particularly apt to such **augmentation**, since they intrinsically provide a live video stream of the operating site and make use of instruments that may be tracked, from said video images or with external tracking systems [92]. The integration of this information with pre- and intraoperative imaging looks like a promising foundation for future developments in advanced visualization, navigation and cognitive assistance. However, several breakthroughs will be required before reaching the accuracy sufficient for meaningful support to the surgeon [11, 173].
Moreover, further incremental and evolutionary work is taking place. **New imaging modalities** and contrast agents being developed, such as the mentioned optoacoustics and multimodal contrast agents, can unlock new approaches to existing procedures, or make new kinds of intervention feasible. The advancement of **visualization** technologies, such as headsets or holographic displays, and hands-free input methods (e.g. gaze or speech) may lead to a more widespread usage of 3D visualization, augmented reality and advanced user interfaces.

3.7 Conclusion

In the previous pages one of the dominating topics was the physical and cognitive **burden** posed by modern surgical setups on the physician, and how it can affect clinical outcome. Unfortunately, this effect is difficult to measure, with the result that it is underrepresented in literature and scarcely addressed in clinical routine. In Chapter 5 we will describe how

introducing one more device into the OR (i.e. a robot) actually simplified the surgeon's task, by taking another apparatus from their hands and performing image fusion for them. Later, in Chapter 6 we will attempt to improve the quality of freehand 3D ultrasound images acquired with the most usable tracking technology (i.e. electromagnetic), so that the operator is less constrained in its usage and can read the resulting volumes more easily.

4 Medical Robotics

In Section 3.3.3 we already introduced robotics as a mean to interact with the real world making use of abstract, digitalized information, and we anticipated a few major application areas. Here we will provide an overview of the field; for the sake of brevity, we will not reach the level of detail offered by notable reviews of the topic [36, 181, 182, 196].

4.1 Motivation

The objective of introducing a new technology into the medical field is always improving the efficacy and/or efficiency of patient care. The adoption of machines in the OR allows to exploit their strengths, which can compensate human weaknesses. Many authors [113, 135, 182] find consensus enumerating these aspects, and in this chapter we will discuss a possible synthesis.

The following advantages can be attributed to medical robots:

Movement accuracy Machines offer extremely accurate movement, exceeding human capabilities. Even though skilled surgeons can reach a high level of hand steadiness, it cannot compete with the stability of a mechanical device. The greatest human limitation in this regard is absolute positioning, i.e. the ability to move to a location specified abstractly (such as a set of coordinates with respect to a reference frame) without a visible reference in its proximity.

Scalability Robots can be designed to satisfy a wide range of constraints in strength, size and shape. This makes them suitable to adaptation for tasks involving the manipulation of objects of very different scale.

Endurance Human performance degrades in time as the operator gets tired. This is not the case for a (well designed) robot.

Resistance to harmful conditions A mechanical device is immune to agents that would be dangerous for a human being, such as high temperatures or exposure to ionizing radiation. This can be advantageous for applications like X-ray or Nuclear Imaging guided interventions or Radiotherapy.

Conversely, the following averse factors must be considered when designing and applying robotic approaches:

Dexterity The human arm is an engineering masterpiece which our technology has still not be able to reproduce, let alone surpass. It is capable of a very high number of degrees of freedom while being able to sustain itself, and even to exercise considerable force. Contemporary robotic devices not only fail to achieve the same performance, but solutions with the same workspace (such as lightweight manipulators) are also usually bulkier and heavier. In order to overcome these limitations it is necessary to design dedicated robotic devices for each application, while this variability would in most cases not constitute *per se* an obstacle for human capabilities.

Flexibility Machines must be controlled by appropriate software. Contemporary systems must be developed specifically for each procedure, foreseeing possible complication and all corner cases in order to guarantee reliability and safety. Changing the behavior of such platform is an expensive process, which must be followed by further approval and certification for medical uses. By contrast, a human can continuously refine their skills from experience, or learn a completely new approach with a comparatively low effort.

Judgement The control of mechanical devices is usually expressed as an ensemble of algorithms and control laws, purposed for accomplishing a task based on the available input. Such input is typically limited to the data provided by the few available sensors and pre- as well as intraoperative images. Employing *closed-loop* control to make use of available data to monitor the process of the intervention and eventually adjust or abort the procedure would be recommendable, but this does not appear to be a universal choice and *open-loop* applications have found their way into the OR. But even when making use of all data available, human contextual awareness and judgement still seem to be out of the reach of contemporary artificial intelligence, as we discussed in Section 3.3.4.

In the light of this reasoning, approaches that use robots as an extension of the surgeon's body to exploit mechanical precision and human intelligence appear to be optimal, at least given the current state of technology. The surgeon can specify the action to be taken via explicit control, as in *human-in-the-loop* scenarios, or declaratively specify a plan which can then be performed autonomously by the machine under strict human supervision. As we will see in the next section, the borders between these fields are blurring over time.

In general, the biggest opportunities opened up by applying robotic techniques in medical interventions are related to the superior mechanical precision, the use of imaging data as reference for motion, remote action or motion filtering via teleoperation, and the possible miniaturization.

Robotic **precision** can be of great advantage when operating on sites cluttered with sensible structures, as during neurosurgery or in the field of urology. Minimizing collateral damage can have great benefit on the long-term quality of life of the patient, especially if nerves can be spared [36]. Limiting the risk or extent of vessel resection can also result in a smaller blood loss during the intervention, giving the surgeon more time to accomplish the task and in turn reducing the stress involved. Alternatively, the robot can act as a non-transparent proxy: the input motion of the surgeon's hands can be modified by scaling them down to manipulate minute structures, or tremor can be removed.

But even more radical processing of the input motion can provide precious support in extreme

cases, such as in **single-port surgery**. In traditional laparoscopic surgery, the tools are inserted through multiple small incisions placed a distance sufficient to allow dexterous manipulation. In the case of single-port surgery, the endoscope and the laparoscopic instruments must fit into a single incision. Since conventional straight instruments could not maneuver under the constraint of the trocar, curved instruments were developed for this purpose. The drawback is that the surgeon must now work in a mirrored fashion, since he/she handles the left instrument with the right hand and viceversa. Such procedure, often called Laparoendoscopic Single-Site Surgery (LESS), turns out to be very challenging to learn and master [83]. Intuitive Surgical developed a Single-Site version of the daVinci robot, capable to swap the movement between the instruments, such that the surgeon can intuitively work as in a conventional robotic surgical procedure [126].

Another possible application are **virtual fixtures** (also called *active constraints*). After patient-to-image registration and robot-to-image calibration, a robot can be programmed to act as a passive guide in order to keep a surgical instrument attached to it, within or outside a region of interest specified on the image itself. This again allows to minimize damage and potentially to perform surgery faster.
teleoperation is usually associated with action over a physical distance between surgeon and patient, but it can be beneficial within the same room as well, as is showed by robotic laparoscopic systems. As we showed in Section 1.2.3, in this case the surgeon sits at a "master" console controlling the robot "slave", which holds the surgical instruments. This setup represents a much more ergonomic configuration than standard laparoscopic surgery; not only the surgeon's health and performance benefit from this fact, but also the patient. While offering a natural user interface to the operator, the robotic slave can perform movements in a cluttered environment which would be very difficult or even impossible for a human, with the result of enabling substantial invasivity and long-term morbidity reductions.
Miniaturization is another factor that allows to minimize collateral damage during surgery, and enables totally new interventions not possible otherwise.

The implications of these propositions will become clearer as we go over a few notable examples of robotic applications in surgery.

4.2 Applications

The major advantage of autonomous and image-guided medical robotic solutions is the geometric accuracy. Making use of this feature is easier when dealing with rigid structures in the human body, such as bones. This explains the wider success of robotic applications in orthopedics [7, 96] and in neurosurgery [113], where *brainshift* can be contained or accounted for in many cases. In the field of general surgery, where unpredictably deformable soft tissues abound, *human-in-the-loop* approaches are the most popular.
In this section we will discuss medical robotic applications according to their degree of autonomy. For complete review, we suggest more exhaustive review papers of the field as a whole [7, 181, 182], as well as of neurosurgical applications [113] and the approaches in general surgery [60].

4.2.1 Autonomous

In Section 3.3.3 we already discussed the issues related to the adoption of completely automated surgical procedures in terms of the relationship between risk, cost, and benefits. Their evaluation must include ethical, legal and economical aspects while ensuring the patient's health. Despite these considerations, autonomous approaches can be found in the literature and have been applied in medical routine. In most cases, they require the surgeon to specify the required action in a plan based on preoperative medical images (often 3D Computed Tomography) which is then executed by the device under passive supervision (as shown in Fig. 4.1).

One of the earliest autonomous surgical systems, called ROBODOC (Curexo Technology Corporation, Fremont, California), is an example of particular interest. Introduced to clinical practice for joint replacement in the early 1990s, it was used on a considerable amount of patients worldwide before controversy arose. A quantitative study of the clinical outcome performed in 2003 [73] highlighted the higher accuracy of the robotic approach with respect to the manual procedure, but did not identify a clinical benefit for the patient in the long run. On the contrary, the disadvantages of the autonomous platform included higher costs and longer average time for the procedure (also due to cases where the system halted and the procedure had to be completed manually). The study also reported a substantially greater magnitude of long-term morbidity and implant dislocation rate, as well as more frequent necessity of a second operation to correct them. As a consequence of the heated public debate and of ongoing lawsuits, the popularity and usage of the system was reduced greatly, in particular in Germany [96]. However, a 2006 ruling of the German high court stated that the post-operative issues were mostly due to faulty surgical plans produced by the surgeon, and not to intrinsic defects of the device. The producer had to shut operations down and transfer the assets to another company, which was able to obtain FDA approval and continue operations outside of Europe. Further studies [4] highlighted the advantages of the approach with respect to the precision of the drilling and the alignment of the implant, as well as the reduced chance of femur fractures during the procedures; both were resulting in a significant reduction in post-operative morbidity.

This case did not stop the research and translation of autonomous surgical systems, but acted as a healthy memorandum for the community. The attention bar for the topics of **benefit/cost evaluation**, the **education** of the operators of autonomous medical devices, the presentation of the **advantages and drawbacks** of such machines to the public, and **ethics** was raised substantially as a result.

Yang *et al.* [205] recently stressed the subject and proposed a regulation scheme with six levels of autonomy. In their work, they highlight the direct repercussions of a medical robotic system performance on the patient's health and draw parallels to the contemporary trend of autonomous driving, which at regime will have to carry a comparable responsibility for an even larger amount of people. They also raise another important topic, which is the lag between the pace of innovation in technology and in legislation regulating their use and responsibility. Another issue they stress is the determination at which a technology is deemed *good enough* for systematic adoption: while achieving perfect performance would be desirable, surpassing human standards could already be sufficient to have a positive impact on global clinical outcome.

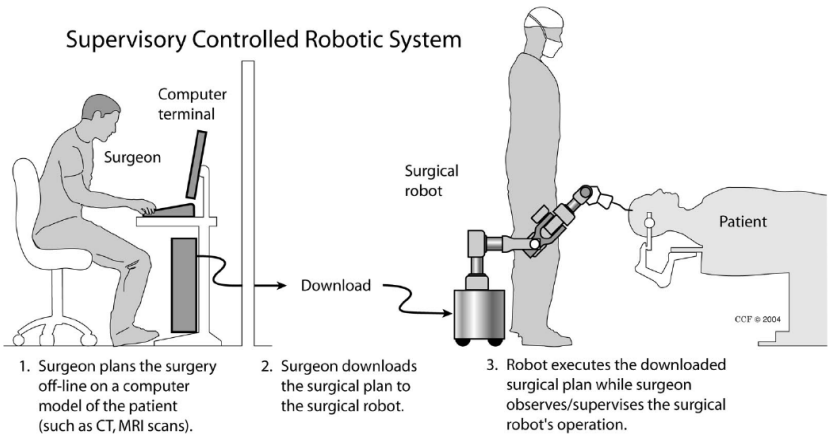


Fig. 4.1. **Control schemes for medical applications: supervised control (autonomous operation).** In this control scheme, the robot operates according to a predefined plan, usually specified on preoperative diagnostic images that are intraoperatively registered to the patient. The surgeon must then pay attention to the progress of the procedure and be ready to abort at the first signals of a deviation from the expected operation.
Reprinted from [135] with permission by Oxford University Press.

In light of the risks associated to automated surgical systems, an easier path is offered by non-invasive robotic applications, such as imaging. The integration of a robotic system for holding an interventional CBCT C-arm has already become commercially available (Artis Zeego, Siemens Healthcare Gmbh, Erlangen, Germany) in order to allow for greater flexibility in the positioning and streamline the surgical workflow with respect to solutions anchored to rails on the ceiling [15, 29]. In general, imaging methodologies do not require invasive manipulation of tissues. While ultrasound imaging requires simple skin contact, other imaging techniques do not require contact with the body at all, since electromagnetic fields and radiation can be detected at a distance. Autonomous Medical Robotic Imaging was showed to be suitable for ultrasound-based automated diagnostic screening [195] as well as for intraoperative multimodal imaging [53]. Our first contribution, presented in Chapter 5 can be classified into the latter category.

4.2.2 Extending

The class of devices which delegate control to the operator is probably the one finding most widespread application. The acceptance seems to be higher, since this kind of architecture allows to solve acute problems in certain fields while leaving the surgeon in complete control of the procedure.
Physiological hand tremor can be harmful to the outcome of many surgical procedures, but in particular in the context of microsurgery, see Section 1.2.4. Stress and fatigue accumulating over the time of the operation have been shown to amplify the problem [66]. As a workaround, approaches have been proposed [176, 183] in order to subtract such tremor from hand movement while maintaining a natural experience for the operator.
A notable use case for robotic techniques is neurosurgery, where the utmost care must is needed in order to avoid damage to sensible structures inside the brain, which could result into irreversible impairment to the patient [113, 135]. Computer-assisted navigation and

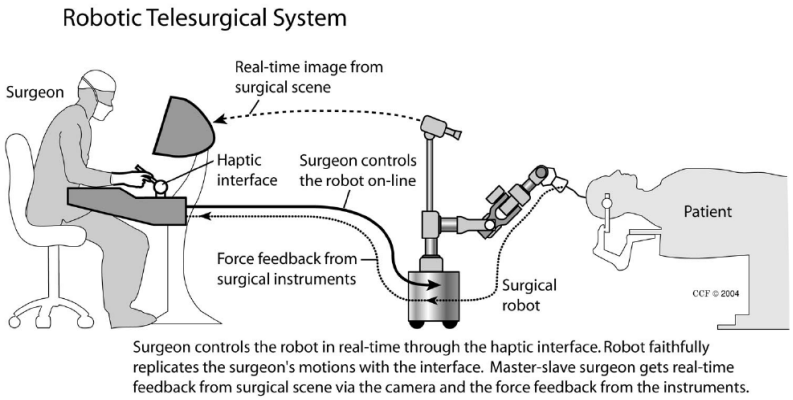


Fig. 4.2. **Control schemes for medical applications: teleoperation.** This approach is currently popular for robotic laparoscopic surgery, since the *human-in-the-loop* scheme allows to employ human intelligence to perform control of flexible instruments in a highly dynamic environment, while taking advantage of motion scaling and tremor filtering enabled by the hardware.
Reprinted from [135] with permission by Oxford University Press.

guidance with visual or auditory feedback can assist the surgeon in avoiding predefined areas, but they do not help with hand tremor or accidental movements. On the other hand, classical stereotaxis can be employed to provide a guide for a needle based on preoperative images, but it does not reach the flexibility offered by robotic applications. In fact, a robot can be programmed to act as a stereotactic guide, i.e. to reach a specified position with respect to the anatomy based on a registered pre- or intraoperative image, and then lock in place. Another important application is limiting the movement of the surgical instrument to a preplanned trajectory in space, or outside of a determined volume (sometimes alternatively defined as *no fly zone*). Such solutions are capable of preventing accidental damage, thus reducing the stress posed on the operator and allowing him/her to focus on the true surgical objective. Similar schemes can be in principle applied to many other classes of procedures, but it is harder to prove benefits capable to outweigh the costs when operating on less sensible regions than the brain or the eye.

In general surgery, the da Vinci family of products (Intuitive Surgical, Sunnyvale, California, USA) is currently the most commercially successful medical platform employing a *human-in-the-loop* approach for robotic laparoscopic procedures [60], (shown in Fig. 4.2). Originally developed for thoracic and cardiac applications, its feasibility for a wide spectrum of procedures has been showed over time [56]. As discussed in Section 1.2.3, the paradigm benefits the patient only indirectly, by providing the surgeon with a better ergonomy and hence improving the accuracy of existing routine procedures [10, 57, 99], or enabling new minimally-invasive treatments, as in thoracic surgery [90]. Unfortunately, such indirect benefits are more difficult to demonstrate. However, even the most skeptical voices criticizing the current benefit/cost proportion stress the enormous opportunities provided by a surgical robotic platform [47, 48]. A breakthrough that could be obtained via an incremental evolution of current robotic laparoscopic platforms would be a further minimization of invasivity, reducing the number of skin incisions required to zero. This can be achieved through approaches that exclusively make use of natural orifices of the body, such as NOTES (Natural Orifice Transluminal Endoscopic Surgery) [157]. Such approaches could further reduce morbidity associated to interventions and recovery times.

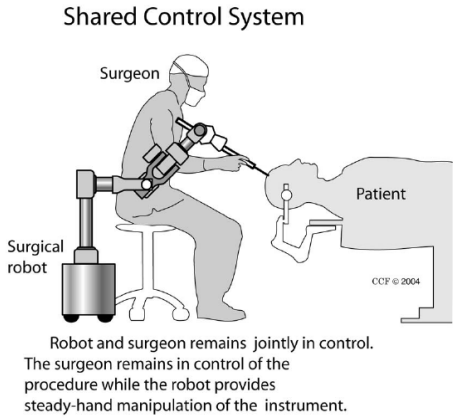


Fig. 4.3. **Control schemes for medical applications: collaboration.** A disadvantage of teleoperation is the distance between the operator and the surgical site. Collaborative schemes allow to overcome this disadvantage while retaining some features such as tremor reduction or virtual fixtures.
Reprinted from [135] with permission by Oxford University Press.

One additional possibility, which we already mentioned in Section 3.3.4, is related to the progressive digitalization of surgery, of which a robotic platform would constitute a cornerstone. The progressive integration of sensors and actuators involved in the surgical process can be envisioned to bring unity to the current patchwork of technologies and methodologies that can be encountered in contemporary ORs, preparing the ground to the introduction of cognitive techniques capable of holistically optimizing the surgical process as a whole and adapting it to the single patient and situations arising.

4.2.3 Collaborating

In the previous sections of this chapter we went over approaches that aim at relieving the surgeon of the manual activities involved in surgery, in order to achieve superior performance. Then we discussed solutions that leave the surgeon in complete control, but try to enhance their capabilities by constraining motion or by acting as a (roughly) transparent proxy. Some of these methodologies have been defined "cooperative", and as such could fit in this section about collaborative robotic applications; an example of this concept is outlined in Fig. 4.3. Here we will focus instead on robots working together with a human, collaborating to achieve the same task by executing a "background" process rather than taking active part in the surgeon's activity.

Implementing a collaborative control scheme is not a trivial task. Achieving an effective collaboration between two or more agents requires a mean of communication. Communication can be explicit or implicit. **Explicit communication** requires an effort on one side to elaborate a representation of the content which is the object of communication into a form which is suitable for transmission; an example of this process is formulating an idea or assembly of ideas into words and sentences. The transmission of said representation typically happens by manipulating a medium in a fashion that can be detected on the receiving side. The most common human mean to achieve that is talking and listening. The received representation must then be decoded back into ideas and concepts, with the hope that they will be understood

as intended by the sender. The success of the process depends on a large number of factors: the ability of the sender to choose an appropriate representation for the content and transmit it as intended to the communication medium, the degradation of the signal while traveling through it, the detection and abstraction capability of the receiver, and the contextual knowledge shared between the entities taking part in the communication process. Any of the aforementioned activities can represent a *single point of failure*; this leads to the installment of communication protocols in most critical environments (an evocative example being the military). The studies we mentioned in Section 3.1 correlating stress and performance also highlighted the importance of communication within an OR for the outcome of the procedures.
Implicit communication can be an effective instrument to replace or complement explicit communication. As social beings, humans acquired notable capabilities to understand a peer's intention by forecasting the development of an ongoing activity or through external factors, such as eye gaze [34]. It is hence in many instances possible to provide assistance in the absence of explicit indications of the next required step. However, such process is even more delicate than responding according to explicit (such as verbal) communication: the interpretation of someone's actions relies heavily on shared knowledge of the ongoing activity, as well as collateral and contextual knowledge ("common sense") [165].

As a consequence, a machine striving at collaboration with a human must be capable of explicit or implicit communication (at least at the minimum level necessary in order to have a positive impact on the outcome of the process). This requires **sensing** capable of detecting signals of communication (e.g. voice commands or eye gaze), as well as the knowledge apt to correctly **interpret and infer** the most appropriate action to be taken next. Not to be underestimated is the necessity for a **feedback** mechanism in order to keep communication alive and notify potential problems. Studies show [37, 50] that collaboration between humans and robots can be improved by optimizing robot movement not for efficiency in accomplishing its own task, but rather for **legibility** of intent by the human counterpart.
The complexity of these tasks explains why collaborative approaches have been the least common in literature and on the market until recent times. However, as we already mentioned, recent advancements in the field of machine perception promise to lead to substantial advancements in Human-Machine Interaction (HMI). As it is possible to detect and recognize the operator's actions and intentions more precisely, it will be possible to exploit this knowledge to automate well-defined tasks in controlled environments and situations through explicitly programmed routines. Future advancements in AI reaching sophisticated contextual knowledge and reasoning will one day allow to scale up and generalize collaborative approaches.

Currently, the most active research in this direction has revolved about automated **endoscope holders**. In Section 1.2.2 we stressed the difficulty of operating the endoscope and the surgical instruments at the same time. One possibility to simplify the surgeon's workflow in this regard is to offer a solution which does not require the use of a hand or error-prone communication with an assistant. The first devices proposed were controlled via a joystick: the surgeon could adjust the position of the camera in order to reach the desired view, and the mechanism would keep it still. Such approaches did indeed avoid potentially ambiguous communication and some guesswork, but introduced the necessity of constantly switching between holding an instrument and the endoscope holder joystick. Later attempts involve collaborative voice-control schemes [166, 167], which result in a hands-free control but are vulnerable to auditory noise in a cluttered environment such as the OR, and pose challenges

with regards to precision. A further alternative can be found in systems controlled through eye gaze [39], which rely on a completely implicit control requiring no action in the typical case. Such methods will surely be able to make use of already mentioned recent developments in computer vision and surgical activity recognition, as well as machine learning approaches to find the optimal assistance strategy.

Furthermore, the feasibility of shared control for **Subtask Automation** has been showed, along with its potential to improve overall clinical performance [9]. While relying on current artificial intelligence for the automation of the whole surgery still poses reliability, ethic and legal problems, the implementation of targeted and well defined routines in an automatic fashion under constant human supervision seems more likely to be viable. This can be of particular utility for simple and repetitive tasks, which are time-consuming for a human operator because of the lack of dexterity of the robotic slave and/or insufficiency in the provided visualization or user interface. In such cases, an *ad-hoc* software can, after explicit command from the user, temporarily take over control of the robot effectors and complete the task autonomously. Research in this direction has been inspired by some of NASA's NEMO programs [63] aimed at compensating for lag while performing telerobotic maintenance in space from a control center situated on Earth. However, even in the absence of lag, such techniques can be useful in the robotic laparoscopic setup for the aforementioned reasons and lead to 10x improvements in speed of execution while maintaining comparable, if not superior, precision in activities like suturing [9].

Approaches making use of pure **Collaborative Robotic Imaging** represent a minority of the field. However, a fitting example can be found in a system for Intraoperative Transrectal Ultrasound (TR-US) Guidance for Laparoscopic Prostatectomy [124]. In this work, the TR-US probe was co-calibrated with a robotic laparoscopic surgical platform in order to keep the TR-US probe pointing in the direction of the active surgical instrument controlled by the operator. Such approach allows an implicit control of the imaging device, making useful information available to the surgeon in real time without explicit communication or coordination efforts involved.
Other studies find practical uses for collaborative robotics outside of laparoscopic procedures. A system developed in the IFL lab employed 3D ultrasound-based visual servoing to implement a collaborative robotic assistant capable to position the US probe in a determined configuration with respect to the anatomy. By attaching a calibrated needle guide to the probe, it was possible to adapt the system to perform guidance for Needle Insertion. After prior specification of the needle insertion site and path on a 3D CT or compounded US volume, the robot will move to a position such that a needle pushed into the guide will follow the specified trajectory within the patient's anatomy. The concept was proven to be a suitable replacement for X-Ray guidance in the context of Facet Joint Injection [44].

4.3 Trends and Outlook

The natural technological evolution will probably lead to further **miniaturization**, with a parallel reduction of invasivity and disruption of the traditional surgical workflow. The development of **flexible robotics** [19] could represent the breakthrough needed to lead to a systematic application of surgery via natural orifices of the body [89]. Further down

the road, the development of Microbots [138] would enable treatments on a micro- and cellular level. In the meantime, efforts are currently being spent in order to develop **special-purpose** robots [60], to be mounted on the patient's spine (MARS) [171], beating heart (heartlander) [150] or skull for retinal surgery [134]. Satisfying exceptional constraints is also driving research into unusual territory, such as the development of MRI-compatible pneumatic robots [182].
Along miniaturization and *ad-hoc* development, another notable trend is the refinement of current technology. The addition of **haptic feedback** [141] to existing platforms is deemed necessary, nice to have or superfluous by different authors [28, 56, 113]. Interestingly, a study [201] reported lack of haptic feedback being a problem not only for robotic surgery, but also for conventional laparoscopy; the authors propose to augment laparoscopic instruments with mechanical actuators to provide artificial feedback based on integrated sensing.

A further identifiable trend is the progressive **integration** with other technologies, which has the potential to make the surgical robot the central platform of the whole OR. We already mentioned drop-in ultrasound and gamma probes that the surgeon can grasp and directly use during the operation; it is possible to imagine more modalities taking a drop-in form or being used to perform continuous background monitoring of the operation with collaborative schemes. An advantage of the integration with devices already present in the OR is the shorter time necessary to obtain certification, which is one of the highest barriers to the commercialization of new medical products.
Furthermore, trends such as Surgical Data Science (which we introduced in Section 3.3.4) that aim at integrating the whole OR into a cohesive **smart environment** are likely to employ the robot's hardware to enrich the surgeon's toolbox.

4.4 Conclusion

In this chapter we outlined some of the trends that can be recognized within the field of medical robotics: from autonomous systems, over to teleoperation, finally to collaborative robots. As the interaction modality becomes more sophisticated, it also grows tighter and more intuitive. We adopted the latter scheme for the design of our collaborative imaging system (described in Chapter 5), with the aim of allowing a direct yet seamless control of the robotic device.

Part III

Proposed Methods

5 Enabling Live Intraoperative US-Gamma Imaging with a Collaborative Robotic Assistant

This chapter discusses our first contribution: an intraoperative multimodal real-time imaging system for guidance during needle punch biopsy of the breast for cancer staging.
We propose a collaborative robot which assists the surgeon in the intraoperative acquisition of multimodal imaging. The robot is programmed to keep the gamma camera in front of the ultrasound probe held by the physician, such that their outputs can be directly merged into a single multimodal image. This is achieved through an optical tracking system, which provides the relative position between the US probe and the gamma camera. The relative position is used for both robot control and image fusion. The surgeon can hence perform a needle biopsy with the free hand, under constant guidance of a real-time multimodal image
After discussing the clinical motivation and the alternative solutions that can be found in literature, we will describe the implementation of the system and the experiments performed, in order to show the feasibility and the unique advantages of our method. Our lab trials showed consistent and substantial improvements in the efficiency and accuracy of the procedure when making use of the collaborative imaging system with respect to the manual alternative.

The core contributions presented here are:

- the first collaborative robotic medical imaging scheme adopting a lightweight manipulator,
- a radically innovative **real-time** intraoperative multimodal imaging device conjugating 2D ultrasound and 2D gamma imaging,
- the application of said system to sentinel lymph node core needle biopsy, with unique advantages over the existing methods.

These results have been published at MICCAI 2015 [40] (oral and poster) and in the IJCARS MICCAI 2015 Special Issue [41] (later awarded as IJCARS MICCAI 2015 Special Issue Best Paper). The pictures thereby contained are reprinted with permission from Springer.

5.1 Clinical Motivation

Breast cancer consists in abnormal growth of breast tissue cells. Due to environmental factors and/or genetic predisposition of the subject, the normal lifetime of a group of cells is perturbed.

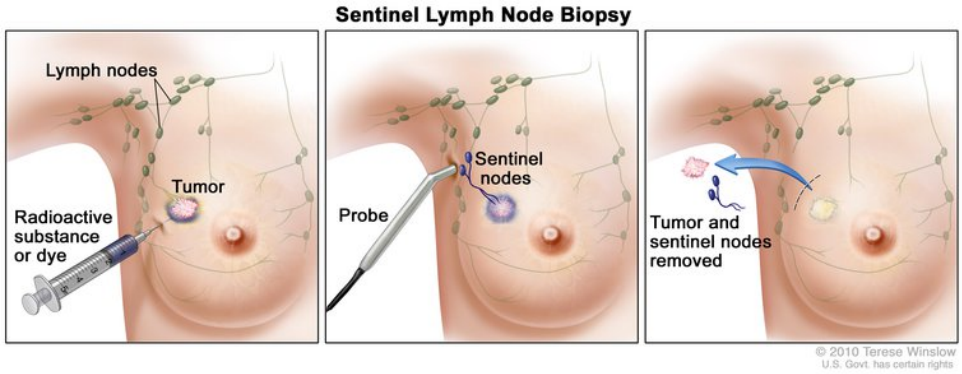


Fig. 5.1. **Sentinel lymph node biopsy in the axilla.** This illustration represents the phases involved in a biopsy of an axillary Sentinel Lymph Node (SLN). In order to find the SLN among the multiple lymph nodes present in the region, it is common to inject a colloid substance in the proximity of the tumor. After a short period of time, the liquid will have migrated to the lymph node (or nodes) that drain the most liquid from the neighborhood of the cancerous mass; as a result, the SLN(s) will be detectable through a gamma probe as well as visible to the naked eye. Typically a conservative surgery follows in order to dissect the tumor and the SLN, which will be later histologically tested for the presence of metastatic cells. Their presence would be a strong indicator that the cancer has started spreading throughout the patient's body.
Reprinted with permission by Terese Winslow.

This can manifest, among other signs, as a change in volume and/or shape of the breast, depending on the depth at which the tumor is located. After a phase of local growth, the cancer can start spreading throughout the body, or *metastasize*. As with most tumors, timely diagnosis and treatment is fundamental to optimize the clinical outcome and minimize the invasivity of intervention.
Breast cancer is overwhelmingly more widespread among females, since they on average have a larger number of breast tissue cells. According to the American Cancer Society [172], 252,710 out of 255,180 new cases of breast cancer estimated in the United States in 2017 affected women. In the same year, breast cancer was responsible for 41,070 deaths among both genders. 5-Year survival rates varied from a 99% for localized cancer to 85% for regional tumor growth, only to sink down to 26% after beginning of metastasis. This abrupt change in outlook stresses the need for accurate *staging* of the disease.

Surgery is the primary instrument for breast cancer management. Depending on the size of the mass, the objective of the intervention may be to remove it (*lumpectomy*), or to amputate the whole breast in a *mastectomy* operation. The former option is preferred and progressively adopted more frequently, but the latter may still be necessary today if multiple lumps are detected or if their size is too large.
In the context of mastectomy or lump excision it was common to remove all the lymph nodes in the area, which usually can be counted in the dozens, with the purpose of subsequent histopathological examination in the search of metastasis. This procedure, often called Axillary Lymph Node Dissection (ALND), is associated with elevated morbidity and permanent discomfort to the patient. In particular, the removal of the lymph nodes leads to a compromised capability of the lymphatic system in the area, with consequent fluid retention (*lymphedema*). It has been shown [193] that the detection of tumoral cells in the closest lymph node to the cancer mass is a key predictor of metastasis, and can replace systematic lymph node dissection.

As a consequence, the process of sampling the tissue of this particular nodule is often referred to as Sentinel Lymph Node Biopsy (SLNB, in Fig. 5.1), in order to highlight its predictive value. This option greatly reduces the invasivity of therapy.
Currently, the clinical routine involves the dissection of said SLN during lumpectomy or mastectomy for histopathological analysis [88]. One of the major challenges posed by this task is to correctly identify the SLN among all lymph nodes (which, as we mentioned, can be counted to dozens in the axillary region): sampling the wrong LN can lead to a false negative, and consequently to a false diagnosis. The current practice employs radioguidance or dye for SLN identification: a contrast medium is injected in the vicinity of the tumoral mass. Over time said fluid will collect in the closest lymph node(s) and make them visible to the naked eye due to its color (usually blue for improved contrast) or fluorescence to a specific electromagnetic wave length. Alternatively, its radioactivity will make it detectable to a dedicated device such as a hand-held gamma camera. The surgeon can then locate the LN visually or by scanning the area with the mentioned imaging device.

While progress in medical routine has already reduced the intervention's invasivity and related morbidity, there is still room for improvement. It has been demonstrated [194] that lymph node dissection during open surgery could be replaced with a less invasive approach such as core needle biopsy (as described in Section 1.2.5), while retaining a false negative rate lower than 5%. After a reliable staging, the lumpectomy could be then performed less invasively without "hunting" for the SLN. Subsequent Radio- or Chemotherapy may be necessary according to the tumor stage and classification.
The major barrier to the adoption of SLNB in place of ALND is a lack of appropriate imaging [82]. While lymph node needle biopsy under ultrasound guidance is feasible, the presence of multiple other LNs in the area makes the intervention prone to error in the identification of the SLN. While employing a contrast medium as in the case of dissection would enable functional imaging capable of unequivocally marking the SLN, functional information alone is insufficient to provide guidance to the surgeon and must be merged with an anatomical modality, such as the mentioned US.
One further source of difficulty is the visibility of LN in US images: many authors [17, 145, 180] report a threshold of 5 mm for the smallest LN visible through US imaging. While it was common opinion that enlarged LN would be more likely to be affected by malignancy, a dedicated study [140] found very weak correlation between LN size and likelihood of metastasis. The same study stressed the fact that LN below the said cutoff size of 5 mm still had a 10% chance of malignancy. Since this means that if the SLN were to be under this threshold it could be invisible to US alone, this would make simple US navigation insufficient for SLNB. However, a multimodal imaging system could help the surgeon spot the subtle profile of a small LN in the US image.
Research in multi-modal imaging for ALND and SLNB is abundant [18]; we will provide an overview of the field in the next section.

5.2 State of the art

The methodologies under development and clinical testing are summarized in some review papers [18, 82].

The current gold standard is dissection during open surgery. Most approaches use multiple redundant factors to reduce the incidence of false negatives. It is common to use a blue dye visible to the naked eye (e.g. Patent V Blue) in conjunction to a fluorescent agent or a radioactive tracer.
Fluorescent materials are chosen to be particularly sensitive to light in ranges where the absorption and scattering in human tissues are minimal, such as the 700-900 nm wavelength range. This allows for an improved contrast against background and deeper penetration. However, the maximum penetration depth remains limited to less than 2 cm.
On the other hand, commonly used **Radiotracers** such as Technetium (^{99m}Tc) emit high-energy gamma particles, which are less sensible to scattering and absorption than electromagnetic waves in the visible range. A dedicated device such as a gamma probe or camera can locate the source of radioactivity with a good directional precision, but it cannot pinpoint the source position along that line.
These factors make radioguidance and Near-Infrared (NIR) fluorescence a **complementary pair**. It is common practice to inject fluorescent agents and radiotracers jointly, so that the location of the SLN can roughly be determined by scanning with a gamma detector and then refined visually by looking for fluorescence. The low tissue penetration of fluorescence makes this operation still invasive, since the surgeon must follow the vague indications of the gamma device on a trial-and-error basis until fluorescence is visible.

Many different imaging devices have undergone development or are still in a research stage. Gamma cameras, capable of providing feedback about the spatial distribution of the radiotracer as opposed to gamma probes, have been proven to improve accuracy for difficult situations, such as when the SLN is close to the injection site. Some models have been equipped with a laser guide in order to pinpoint the center of their field of view on the tissue surface.
A hybrid gamma camera, composed of a gamma camera and an optical RGB camera, has also been developed in order to match fluorescence and gamma information in a single multimodal image. This was achieved with a semi-transparent mirror placed in front of the gamma camera collimator. The aim is to solve the logistical challenge of using two separate imaging systems and reduce the burden of mentally matching the two outputs, which affect other approaches.
Other research groups are investigating the adoption of alternative physical principles. Gamma-Beta imaging employs two different radiotracers, with the latter capable of a higher space resolution to the cost of lesser tissue penetration. Beta-OCT tries to provide anatomical information via optical coherence tomography, which is unfortunately very limited in tissue penetration as well. Gamma-MR attempts the same, but the characteristics of magnetic resonance imaging make it challenging to adopt in an intraoperative scenario.

The aforementioned methods are only suitable for dissection in the context of open surgery, due to their limited tissue penetration depth. However, there are some efforts to develop images capable to support SLNB.
Freehand SPECT (in Fig. 5.2) was developed as an aid for lymph node biopsy. As mentioned in Section 2.2.4, it employs a tracking system to follow the position of a gamma camera as it is used by an operator to manually scan a region of interest (ROI). The result is a volume where the intensity of the radioactivity originating in each voxel is estimated, and can be used for interactive visualization during the intervention. If an ultrasound probe is simultaneously tracked by the same system, it becomes feasible to dynamically reslice the volume to obtain a functional image co-located with the respective US image, and the two data sources can

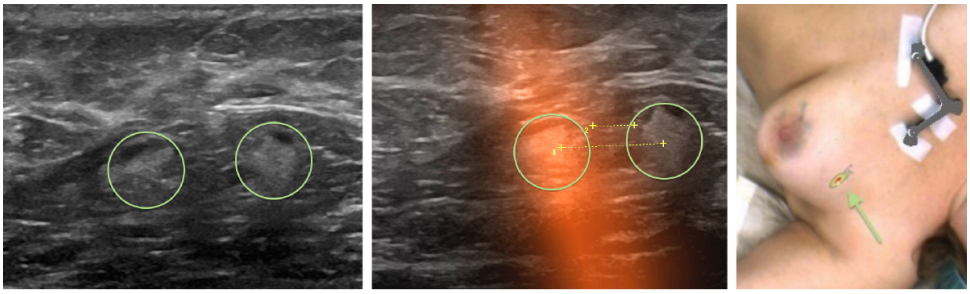


Fig. 5.2. **Freehand SPECT for SLN biopsy in the axilla.** On the left: an US image depicting two LNs (highlighted in green). In the center: an overlay between an US image and the respective cross section of the freehand SPECT volume. The additional functional information added to original image allows to infer that the left LN is marked with the radiotracer, and hence is the SLN. While the US image updates in real time, the same can not be said for the SPECT volume which is acquired intraoperatively just before the beginning of the biopsy. On the right: the operating site of a freehand SPECT and US guided SLN needle biopsy. The optical marker visible on the chest of the patient is used to maintain the alignment between the anatomy and the acquired volume; but it cannot help against eventual deformations of the soft tissue of the breast.
Image from [142], reprinted with permission from Springer.

be merged. With this approach, intraoperative multimodal imaging can be obtained for depths sufficient to perform a SLNB. The major issue with this method is that the freehand SPECT acquisition must be performed in a phase antecedent to the actual biopsy. Any patient movement or tissue deformation would lead to misalignment between the functional and anatomical images. This is a particularly big problem for the region of the female breast, which is largely composed of soft tissue; hence the system may be more apt to less deformable areas such as the neck. Moreover, the operator's skill in covering the ROI has a big impact on the quality of the final reconstruction.
The same considerations about tissue deformations hinder the usage of preoperative images, such as SPECT/CT, PET/CT or PET/MR.
A further attempt at multimodal imaging is the ECORAD project [146], which has the aim of developing a hybrid US-Gamma device. While such a device would allow real-time multimodal imaging, it has not yet become commercially available. The last published results consist of promising Monte Carlo simulations and phantom experiments.
A further possibility involves the usage of ultrasound contrast agents, such as micro-bubbles. The last available tests report an insufficient sensitivity [31, 153].

In the next sections we will describe our proposed system, which allows to overcome the limitations of traditional approaches in terms of tissue penetration as well as real-time imaging.

5.3 Technologies

The system that we propose is large and complex. It is worth to take a brief look at the fundamental building blocks before delving into the overall system architecture.

5.3.1 Ultrasound B-mode

As we mentioned in Section 2.2.3, ultrasound imaging is based on the reflection and refraction of high-frequency mechanical waves in tissues. The waves can be generated and then detected by the same transducer array, composed of elements capable of converting electrical energy into mechanical, and viceversa (this phenomenon is known as the *piezoelectric* effect). While a large number of modalities have been developed that are able to image the distribution of different tissue characteristics and physiological processes, we will focus on the most common option employed to obtain understandable pictures of internal anatomy: B-mode (also known as 2D-mode).
In B-mode, a linear transducer array is excited with a high-frequency pattern in order to create an arc-shaped wave. This wave travels through the surface of the US probe (typically a plastic or rubber material to facilitate transmission) and a water-based gel (applied to the skin of the patient to improve acoustic coupling) into the tissue of interest. While traveling through tissue, the wave will be absorbed, scattered, reflected and/or refracted, according to the variations in density of the tissue layers. Depending on the time of flight and the intensity of the echo reaching back to each transducer array element, it is possible to organize this input into a regular pixel grid to form a grayscale image.

While being a harmless, fast, cheap and flexible modality, ultrasound imaging is not without limitations. In its common 2D form, it is only possible to image single planar views of an area; mastering the art of finding an object of interest and reaching the desired point of view requires a high level of skill and places a heavy burden on the operator. Furthermore, the image usually offers a low Signal-to-Noise Ratio (SNR) due to the abundant reflection and scatter. Finally, the quality typically degrades along the depth dimension, since the wave attenuation increases exponentially with the traveled distance [91]; shadowing is also a possibility if a very dense object (such as bone) finds itself between the probe and the object of interest.

More details about B-mode beamforming and image formation process are available in textbooks regarding the subject [74].

5.3.2 Portable Gamma Cameras

Gamma cameras have already been mentioned in the previous sections (2.2.4 and 5.2) for scintigraphy and freehand SPECT. For our project we used a portable handheld gamma camera, which was constructed to employ the same physical principle of conventional gamma detectors. Most of them do indeed exploit *scintillation* to detect when a particle originating from radioactive material enters its field of view and interacts with the matter in its *scintillation chamber*. This occurrence is defined as *event*.
In order to detect and count these events, a composite system is used. A flat crystal capable of interacting with the traveling gamma photons is placed in front of an array of photomultiplier tubes. When a gamma photon interacts with the crystal, it gets absorbed by an atom of the crystal; as a result, an electron is raised to an excited state. When the electron returns to a lower-energy state, the residual energy is emitted again as another photon of a characteristic wave length, typically in the visible range. As a result, the photomultiplier tube is able to

detect this flash and convert it in an electric signal which can be digitized.
A single chamber can detect the incoming photons; the number of events detected per second can be represented with a visible or audible signal. However, by juxtaposing multiple chambers into a 2D array it is possible to employ the information provided by each of them to compose a 2D grayscale image.

In order to enhance the spatial resolution of the gamma camera, it is necessary to minimize the angle at which an incident photon can enter the scintillation chamber. This can be done by placing a metallic regular grid in front of the detector. The ratio between the size of each cell and the length of the *collimator* determines the degree of filtering of incoming particles. This ratio is subject to a trade-off: a longer collimator will decrease the spurious particles reaching an element of the array and consequently image blur, but this will also turn in overall lower counts and possibly SNR; to the limit, this could result in the inability to detect significant information.

5.3.3 Optical Stereo Tracking

Calibrated video cameras have now long been used for tracking objects of interest in space, in particular using markers of known geometry and capable of providing high contrast features; QR codes or ArUcO [159] markers can be considered notable examples. Typically black-and-white patterns are used, such that they can be detected with simple thresholding of grayscale images under different lighting conditions. These patterns are usually composed of geometric primitives, such as straight lines or circles, so that they can be detected via computationally cheap edge- or blob-detection and shape-fitting algorithms. The result of these pipelines is the location of the marker outline on the 2D image.
Deriving the 6 degrees of freedom (DoF) position and orientation in 3D space from the 2D outline on the image plane requires a careful modeling of the image creation process. The most common camera model is the so-called *pinhole model*, which can lead to accurate results with standard digital cameras (after appropriate calibration). In this model, the camera has an *optical center*, into which all incoming light rays are made to converge from the camera lens. The rays intercept the camera sensor, posed behind the optical center, and are quantized into a digital signal by its finite elements (*pixels*). Ignoring the final quantization, the image formation process can be described as a **geometric projection** of 3D points on a 2D surface, i.e. the image plane. The representation of a 3D projection can be conveniently expressed by making use of 4D *homogeneous coordinates*, which can be derived for each 3D point by adding a 1 to the original 3D vector:

$$\boldsymbol{X_h} = \begin{bmatrix} \boldsymbol{X} \\ 1 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ 1 \end{bmatrix} \tag{5.1}$$

Where $\boldsymbol{X}$ is the original 3D vector and $\boldsymbol{X_h}$ is the resulting homogeneous vector. This process is known as *homogenization*. It is worth noting that in Projective Geometry homogeneous vectors are equivalent with respect to multiplication; that is, $\boldsymbol{X_h} = m\boldsymbol{X_h}$ where $\boldsymbol{X_h}$ is an homogeneous vector and $m \neq 0$ is an arbitrary factor. In order to avoid confusion, it is a common convention to *normalize* vectors by dividing them by the last (homogeneous) coordinate, such that it is 1. After normalization, it is possible to extract the non-homogeneous vector by omitting the last coordinate.
The projective transformation can be represented by a matrix establishing the relationship between the points in 3D space and the respective projected point on the 2D image plane. Since we adopt homogeneous coordinate systems for both spaces, this results in a 3×4 transformation matrix P:

$$\boldsymbol{x_h} = P\boldsymbol{X_h}. \tag{5.2}$$

Here $\boldsymbol{x_h}$ is the homogeneous coordinate of the projected point on the image plane. Following a popular convention [65], we can decompose P into two matrices:

$$\boldsymbol{x_h} = P\boldsymbol{X_h} = K[R|t]\boldsymbol{X_h} \tag{5.3}$$

where K is the **intrinsic camera matrix**, and $[R|t]$ is composed of the rotation and translation of the camera with respect to the world coordinates. The particular choice of the world coordinates is arbitrary, but determining K is necessary in order to use the images provided by the camera to determine its position in space and that of other objects. This consists in computing the camera's focal length f, the location of the camera's principal point x_0 (corresponding to the intersection of the principal axis and the image plane), the pixel size in each direction m_x and m_y, and the *skew* of the pixel grid and of the lens deformation s:

$$K = \begin{pmatrix} fm_x & s & x_0 \\ 0 & fm_y & y_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{5.4}$$

However, such model is still incomplete for most commercial systems, since it assumes an ideal projection process. This is usually not the case for cameras making use of a lens in contrast to an abstract *camera obscura*. Imperfections in the lens manufacturing process and its assembly within the camera add further disturbances which must be taken into account. In virtue of this, additional **distortion models** have been developed that aim at approximating the effects of the most common lens properties [215]. One is **radial distortion**, which can be

approximated given the knowledge of the radial distortion coefficients k_1 and k_2. Given these coefficients and the non-homogeneous projected point x, the distorted coordinates

$$\boldsymbol{x_d} = \begin{bmatrix} x_d \\ y_d \end{bmatrix}$$

can be computed to compensate both effects as in:

$$\begin{aligned} x_d &= x(1 + k_1 r^2 + k_2 r^4) \\ y_d &= t(1 + k_1 r^2 + k_2 r^4) \end{aligned} \tag{5.5}$$

where $r^2 = x^2 + y^2$.

The determination of the camera parameters is called *intrinsic camera calibration*. Camera calibration is a vast research field, and many calibration methods have been proposed. Most of them employ a *calibration pattern* of known geometry and size, such as a chequered pattern, which must be moved all over the camera FoV. The images of the pattern acquired in such procedure can then be used, together with the knowledge of the focal length of the camera, to constrain the parameters of the pinhole and distortion model into an overdetermined equation system, whose solution corresponds to the elements of the camera calibration matrices.

Once a single camera has been calibrated, it is possible to use it to track said markers of known shape and size in 6 DoF. The accuracy will degrade as the distance from the camera increases, as a shift in one pixel back-projects to a bigger displacement in space. Moreover, the detection can be unstable when a symmetric marker is orthogonal to the camera axis, since the shape fitting can be ambiguous for such configurations.

It is possible to improve the tracking accuracy and stability, to allow the use of simpler markers, or even to enable markerless tracking by making use of multiple cameras concurrently. The direct extension of a single-camera system makes use of two cameras, and is called a *stereo tracking system*.
After intrinsic calibration of each camera, it is possible to *triangulate* the position of the detected features in 3D space between the images provided by each device to refine the detection. However, this requires first the knowledge of the relative position of the two cameras; *stereo calibration* consists in the process of finding this geometric transformation, which can be expressed through the **fundamental matrix** F which correlates the position of corresponding points x and x' in the image of the two cameras:

$$x'^T F x = 0. \tag{5.6}$$

With this knowledge, the objective is then to find the point in 3D space $\boldsymbol{X}$ which gets projected into the two corresponding points $x = P\boldsymbol{X}$ and $x' = P'\boldsymbol{X}$ and satisfies Eq. 5.6. Unfortunately, reality is not as straightforward and this equations will never be satisfied exactly because of

measurement errors. A large number of methods have arisen to find the estimates which minimize the deviation from these constraints [65].

Analogously to the single camera case, most approaches developed to find the fundamental matrix use a calibration pattern or object; some methods can perform intrinsic and stereo calibration from a single acquisition.
With a stereo tracking system it is possible to employ simpler markers, such as geometric primitives, which can be triangulated between the two cameras as mentioned above.

More details about stereo tracking and the geometry of computer vision can be found in literature, for example in [65].

5.3.4 Collaborative Robotics

In Section 3.3.3 we introduced the definition of a robot as a machine capable of exercising mechanical force on the environment in order to modify it or for self motion. This statement is very general and is meant to encompass the large variety of forms and shapes in which robots can be designed in order to perform a specific task in a specific environment.
Here we will go into further detail for a particular class of robotic devices, which have been designed for collaboration with humans in mind: **lightweight manipulators**.

A manipulator, also called a *robotic arm*, is usually a single chain of rigid *links* connected by *joints*. The most common joints are *revolute* joints, i.e. acting like a hinge between the two connected links; however, other types of joint exists.
Each joint is provided with a **motor** in order to be able to change the state of the joint; in case of revolute joints, this is the opening angle between the two connected links. Usually an **encoder** is also built into the joint, so that the current state of the robot joints is known and it is possible to derive the position in space of the whole robot.
The computation of the position of each link in space is denominated **Forward Kinematics** (FK). A prerequisite is to define the coordinate systems that describe the positions and orientations in space. Following a popular convention [32], we will use curly braces to identify reference frames; for example we will assign the reference frame $\{base\}$ to the base of the robot. We will describe points identified by coordinates in this reference frame as

$$ {}^{base}P = \begin{bmatrix} p_x \\ p_y \\ p_z \end{bmatrix} \tag{5.7} $$

where P is the point identified in 3D space by the coordinates p_x, p_y and p_z. We will write a rigid Euclidean transformation, i.e. a 6 DoF translation and rotation, between the two reference frames $\{base\}$ and $\{effector\}$, as ${}^{effector}_{base}T$. For a robot with a single joint connecting the base with a link named $\{link\}$ ending with the end effector, we can write

the transformation between the base and end effector as the composition of the transforms between base and first link, and from the first link to the end effector:

$$^{effector}_{base}T = ^{link}_{base}T \, ^{effector}_{link}T. \tag{5.8}$$

The choice of the location of the reference frames with respect to the robot links is completely arbitrary, but it has a great impact on the FK computation. One popular system to attach the reference frames to the robot link in order to achieve simple and elegant dependencies between links is the Denavit-Hartenberg convention [35]. However, beside not being the only method available, it it also has a range of variants depending on choices left open by the convention itself. Anyway, for this scope it is enough to say that it is usual to assign the reference frame for each link such that it aligns with the joint axis connecting to the previous link. If the reference frames are assigned in this fashion, it is then possible to derive for each link $\{linkN\}$ the transformation $^{linkN}_{linkN-1}T$ between the parent reference frame and its own. By composing a chain of these transformations, it is then possible to derive the full kinematic chain, depending on the state of each joint and the length of each link.
While a manipulator's Forward Kinematics has a closed-form solution which is complicated for a human but straightforward and manageable for a computer, the inverse process is not so in general. Finding the joint states such that the end effector reaches a target 6 DoF position $^{target}_{base}T$ with respect to its base constitutes the problem known as **Inverse Kinematics** (IK). In contrast to FK, the IK of a revolute joints-based robot gives origin to a non-linear set of equations. While there are systems capable of generating an analytical solution for a given robot kinematic model, the most common solution is to adopt iterative numerical solvers. However, there may be no solution to the problem if the target position is outside the **dexterous workspace** of the robot. The dexterous workspace is so called in order to discriminate it from the adjacent region of the *reachable workspace*, which is the set of points that the robot can reach but only for a restricted set of possible orientations of the end effector. On the other hand, the problem may even have *infinite* solutions if the robot is **redundant**, i.e. it has more degrees of freedom than the space it operates in (e.g. 7 links for a robot operating in the 6 DoF 3D space, or 3 links for a robot working on a 2D plane). In this case a solution must be chosen among all possible ones, arbitrarily or according to a secondary cost function added to the solver in order to satisfy some constraint (e.g. obstacle avoidance).
Once the joint configuration leading to the desired target robot position has been computed, it is necessary to determine the full **trajectory** from the starting to the target position. This means to find a set of positions, parametrized over time, within a small threshold from each other, joining the former to the latter. While the problem could be solved by variating the position of the joints from the starting to the desired configuration, the movement being originated in Euclidean space would be very unpredictable: the danger is to hit an object in the environment, or even the robot itself in a *self-collision* event. Since this is in most cases unacceptable, various strategies can be employed to minimize these risks: some robots cannot enter a state of self-collision within their joint ranges by design. Collision with the environment can only be prevented if it is perceived in some way; a possible solution is to employ an RGBD camera to build a 3D map of objects present in the surroundings of the robot. Anyway, even when such a map is available, an analytic solution proves to be intractable; the state of the art is represented by probabilistic algorithms which sample the configuration space of the robot to find a set of collision-free positions to be joined into a smooth trajectory [27].
Another common pitfall for robotic applications are so-called **Singularities**, i.e. joint configura-

tions where the movement of a joint does not contribute to the movement of the end-effector; this phenomenon is related to the well-known *gimbal lock* problem. In such cases the robot movement can reach dangerous speed in order to maintain a constant speed of the last joint, or completely arrest if the robot software does not find a way to go beyond the singular point. Singularity avoidance is a popular secondary objective of many robot control routines [1, 133].

In our overview of the vast field of robotics we have not considered interaction with humans yet. Robot producers started designing robots explicitly for collaborative scenarios, introducing *ad-hoc* control strategies and safety features. The robot that we employed in this work belongs to this category. As most redundant manipulators, its "elbow" can be moved while keeping the end effector still in its place. This feature can be useful for object avoidance, or to make space for a human in a dynamic environment. The geometric configuration of the robot was designed such that self-collision is impossible within the joint range, and no particular care must be taken when commanding positions to the robot. The torque sensors built into the joints allow to estimate the force acting on each joint, so that an emergency stop can be triggered upon collision with an object or person. These features, along with the industrial-certified motion controllers, were helpful in safely performing our experiments.
The embedded torque sensors alone can enable basic interaction with the environment, at least in controlled ones. They allow to perform an emergency stop if the torque applied by a joint surpasses a threshold; this can be a useful safety feature. However, it is often desirable to rather constrain the force applied by the robot end effector, expressed in Cartesian space. This could be achieved by installing a force/torque sensor between the robot flange and the end effector which can directly measure this quantity, but such hardware is expensive. The same functionality can be partially emulated using the data provided embedded torque sensors, given the current position of the robot. This requires computing the robot's **Jacobian** matrix, which expresses the relationship between quantities in the *joint space* of the manipulator (hence expressed as vectors of length l, where l is the number of links of the robot) and quantities in the Cartesian space (expressed as vectors of length 6, equal to the number of degrees of freedom of the Cartesian space). The Jacobian matrix is commonly denominated $J(\Theta)$, where Θ is the vector of the current joint angles, to stress the fact that it is dependent on the position of the robot. While numerical methods exist to derive the Jacobian of an arbitrary robot in a given position, the matrix can be geometrically derived for manipulators, thanks to their particular architecture. Once the Jacobian is known, it is possible to derive the Cartesian velocity of the end effector (in the reference frame of the base of the robot) ν from the current velocity of rotation of the individual joints $\dot{\Theta}$. This relationship can be formulated as:

$$\nu = J(\Theta)\dot{\Theta}, \tag{5.9}$$

where $\nu = [v_x, v_y, v_z, \omega_x, \omega_y, \omega_z]^T$ concatenates Cartesian linear and angular velocities v and ω. The same can be done to find the Cartesian force produced by the torque applied by a motor:

$$\tau^T = F^T J(\Theta), \tag{5.10}$$

where τ is the column vector of length l holding the torques being produced by the robot joints, and F is the column vector of the Cartesian force f and torque n along each axis: $F = [f_x, f_y, f_z, n_x, n_y, n_z]^T$.
Equations 5.9 and 5.10 suggest that it should be possible to compute the joint velocities

required to achieve a target Cartesian velocity by using the inverse of the Jacobian, and that the same could be done to find the Cartesian force/torque exerted by applying specific motor torques. This is indeed possible, except for positions at which the Jacobian matrix does not have full rank and cannot be inverted; these configurations constitute the *singularities* that we mentioned in the previous paragraph. In this cases, naive computation of the required quantities can lead to numerical instability and unbounded solutions. The situation is even more complicated for redundant manipulators, for which the Jacobian matrix is not square. For this class of robots, inverting the aforementioned equations will lead to an undetermined problem, with the consequence that an auxiliary constraint must be added in order to obtain a unique solution and at the same time a predictable behavior (a simplified example would be: the robot "elbow" must be pointing in the same direction).

If not only the geometric configuration of the robot is known, but also the mass and mass distribution of its constituent links, it becomes possible to compute **dynamic** effects. Equation 5.10 allows us to find the motor torque needed to exert a force on the environment. The following equation provides us with the torque necessary to move the robot in space, without exerting such force:

$$\tau = M(\Theta)\ddot{\Theta} + V(\Theta, \dot{\Theta}) + G(\Theta), \tag{5.11}$$

where $M(\Theta)$ is called the $l \times l$ mass matrix of the manipulator, $V(\Theta, \dot{\Theta})$ expresses the centrifugal and Coriolis effects, and $G(\Theta)$ represents the effect of gravity on the robot. The first term of the equation summarizes Newton's second law of motion $F = ma$ for the whole robot in matrix form, so it is non-zero when the acceleration is also non null. The second term is a complex function of the robot position and velocity.
Equation 5.11 is important because it allows us to implement a control mode which is often called **Gravity Compensation**. Applying the motor torques resulting by solving for this equation, the robot does not exert any force on the environment. This means that it will remain still, until something or someone pushes it; for example, a human operator can move the robot freely in space. This modality enables a set of collaborative applications, and we used it extensively in Chapter 6. It is also a useful instrument to implement "teaching" mode; that is, the robot can be moved by the operator in space to reach a series of positions, which are recorded. The robot can be then commanded to reach these positions in sequence indefinitely. Such method is popular for industrial applications, e.g. in assembly lines.

In this chapter we discuss the implementation of a collaborative control scheme. We already discussed more high-level details of the interaction between humans and robots in Section 4.2.3.

5.4 Proposed Method

In Section 5.1 we discussed the difficulty to reliably target the SLN during ultrasound-guided needle biopsy. In light of these considerations, real-time intraoperative multimodal imaging may be helpful to increase the sensitivity of the procedure and make it a reliable tool for breast cancer staging. We propose a system, shown in Fig. 5.3, capable of fulfilling this task by augmenting the standard ultrasound guided needle biopsy workflow with a collaborative

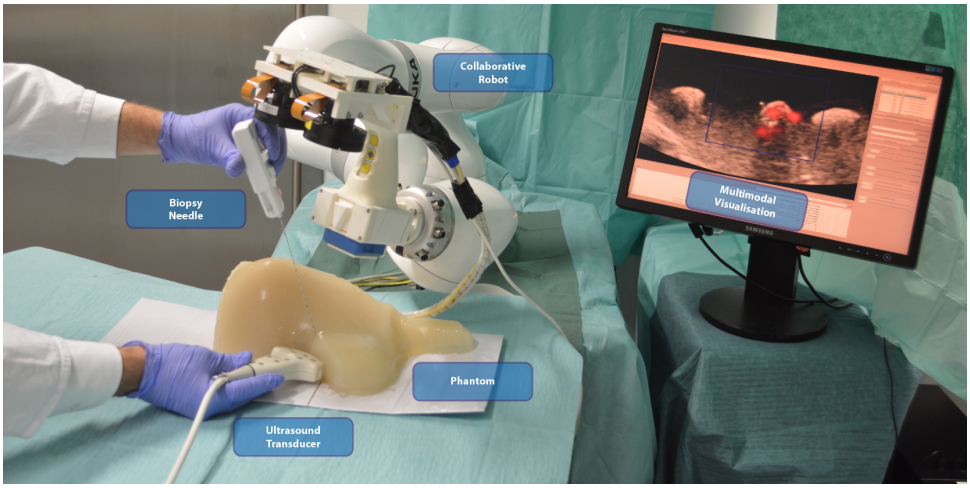


Fig. 5.3. Overview of the Collaborative Robotic ultrasound-gamma imaging system. The situation shown here is an image-guided needle biopsy of sentinel lymph node. The phantom is shaped after the female anatomy of the axilla, and contains clusters of lymph nodes. For each cluster, some LNs are marked with a radiotracer. The screen on the right shows the output of our collaborative multimodal imaging system. A blue square indicates the portion of the US image which is currently within the field of view of the gamma camera. In this region, red dots highlight the subregions which may be the origin of gamma activity. However, the actual location of the radiotracer may be behind or in front of the US imaging plane.

robot which holds a gamma camera. The US and gamma images are then merged in real-time to provide a multimodal image. In this section we illustrate the design of the system.

5.4.1 System Architecture

The system is composed of a network of computers, sensors and effectors. The backbone is formed by the "Ultrasound Workstation" embedded in the ultrasound machine, the "Visualization Workstation" dedicated to robot control and visualization, and one more computer running the optical tracking software ("Tracking Workstation"). A high-level overview is shown in Fig. 5.4.

The ultrasound machine is an UltraSonix RP system (Ultrasonix, MA, USA) running the native software with Research Mode activated and connected to a C5/60 curvilinear probe. Directly running on the same system is also an instance of the Plus [97] toolkit and OpenIGTLink [186] server, which are used to transmit the formed B-mode image and related metadata (such as pixel spacing) retrieved through the Ultrasonix Ulterius SDK.

The tracking workstation is connected to a custom-made optical stereo tracking system over Ethernet, and to the robot's workstation through another Ethernet cable inserted into another port on a PCI-Express Gigabit Ethernet Card. The FRAMOS optical tracking software FIS (FRAMOS GmbH, Germany) runs on the computer, driving the tracking hardware and providing the current position of the marker in real time over an OpenIGTLink communication to the robot control workstation.

The latter is connected via a Gigabit switch to the ultrasound machine and to the tracking workstation on one port, and to the robot cabinet through another Ethernet port. The software running on it includes the visualization software CAMPVis [169] and a collection of processes

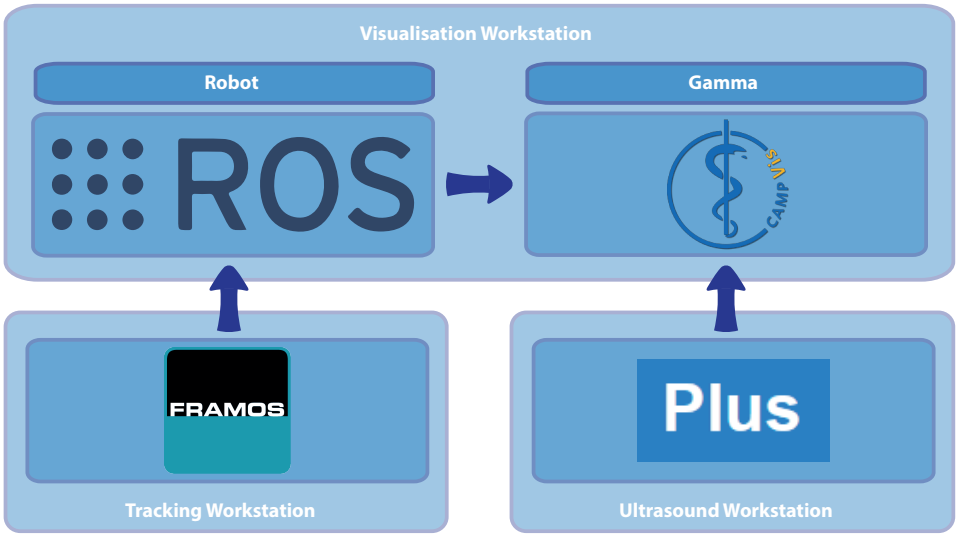


Fig. 5.4. **High-level software components.** Their hardware deployment is also represented in the picture.

implementing the robot control, built on top of ROS [156]. Finally, the KUKA LWR iiwa R800 (KUKA Roboter GmbH, Augsburg, Germany) is connected over a custom connector to its cabinet, hosting an instance of the KUKA Sunrise control software. We developed a library (iiwa_stack [71], see Appendix C) which is used to pilot the robot by updating the commanded position sent over the Ethernet cable from the workstation on a ROS topic. The robot holds a CrystalCam (Crystal Photonics, Germany) gamma camera through a custom designed and 3D printed mount. The same mount supported the stereo camera assembly of the optical tracking system.
The tf ROS subsystem is used to maintain a representation of the current geometric configuration of the whole system. Being a distributed system making use of ROS topics, a custom made bridge between ROS topics and OpenIGTLink connections is used. In particular, it forwarded the spatial properties of the US image as provided by Plus and the output of the optical tracking system to the robot control routines. The CAMPVis instance connected to both the Plus OpenIGTLink server and to the tf ROS subsystem in order to fetch the information necessary for the visualization, i.e. the current US image and relative position between the US probe and the gamma camera.

5.4.2 Conventions and Geometric Reference Frames

The choice of the reference frame attached to notable objects in the system is in most cases arbitrary. Here we will define them in an unambiguous way; Fig. 5.5 contains a simplified scheme of the relationships between the most notable ones.

In this scope we are only interested in the displacement between the robot base and flange ${}^{flange}_{base}T$, and not in the position of the single links of the robot. For the $\{base\}$ and $\{flange\}$ reference frames we use the same convention adopted by the KUKA Sunrise software and iiwa_stack.
The robot end effector consists of a custom 3D printed mount holding the gamma camera

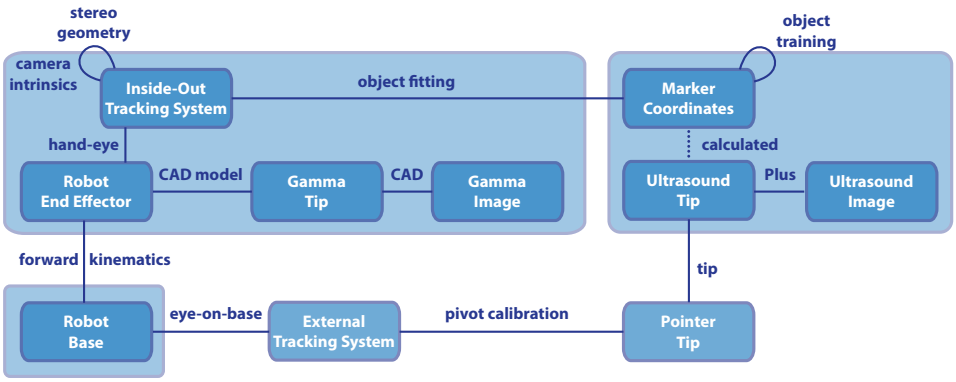


Fig. 5.5. **Notable reference systems.** This scheme summarizes the relationships between the most important reference frames used involved in robot control and visualization.

and the optical tracking system, fixed to the robot flange via screws. On this end effector two reference frames of interest can be identified: one relative to the gamma camera screen $\{gamma\}$, and one to the optical tracking system $\{tracking\}$. The $\{gamma\}$ reference frame is located in the center of the gamma camera screen, with the z axis pointing out and the x axis pointing "up", away from the robot flange. Since we designed the mount holding the gamma camera based on its CAD model, we were able to derive the ${}^{gamma}_{flange}T$ transformation by construction. The $\{tracking\}$ reference frame is arbitrarily determined by the FRAMOS OTS system during the intrinsic calibration procedure. In order to find the ${}^{tracking}_{flange}T$ transformation we relied on a hand-eye calibration procedure, as described in Section 5.4.5.
As we discuss in the following Section 5.4.3, the reference frame attached to the tracking marker is by the optical tracking system software. Since the Plus toolkit is able to provide in real time the geometric transform between the US probe "tip" and the top left corner of the image ${}^{image}_{tip}T$ according to the current parameters set on the US machine, we adopted a custom calibration procedure described in Section 5.4.5 to find the transform ${}^{tip}_{marker}T$. According to the Plus documentation [1], our $\{tip\}$ reference frame corresponds to Plus' *transducer* frame, our $\{marker\}$ frame is analogous to the Plus $\{probe\}$ frame, and $\{image\}$ refers to the top left corner of the US image in both cases.

5.4.3 Optical Tracking

The stereo tracking system is composed of a pair of C1291M-BL boardlevel cameras (SMARTEK Vision, Croatia) with 135°wide-angle lenses DSL315B-NIR (Sunex, USA), mounted on a custom 3D printed support. The lens of each camera is surrounded by a FLDR-i70A direct ringlight (FALCON Illumination, Malaysia), illuminating the scene at a wavelength of 875 nm in the near infrared spectrum. The ringlights are controlled by an IPSC2 (SMARTEK Vision, Croatia) strobe controller. The relevant spectrum for the image processing is selected through IF 093 NIR (Schneider-Kreuznach, Germany) filters.
The tracking target consisted of a set of 10 self-adhesive, retro-reflective circular markers (3M, USA), each with a diameter of 10 mm. The markers are placed on one of the two faces of the ultrasound probe in a highly asymmetric configuration, in order to ensure that the pattern

[1] Plus Documentation: Coordinate Systems Definitions used by fCal (`http://perk-software.cs.queensu.ca/plus/doc/nightly/user/ApplicationfCalCoordinateSystemDefinitions.html`)

would be univocal even in case of partial occlusion.
The tracking algorithm has been previously presented in [21]. After thresholding of the image and detection of the ellipses corresponding to the markers in the image, their centers are computed. Then the PMF Algorithm is employed to find a triangulation of the detected center points and determine their position in 3D space. Finally the pose of the whole probe is determined by matching the points found to the configuration found during a reference acquisition. This acquisition is performed beforehand in a steady state, averaging the pose computed for each point over time to ensure robustness to outliers. After the acquisition, a reference frame is attached to the constellation of the markers in order to represent the 6-DoF position and orientation of the US probe. The choice of the reference frame is completely arbitrary, but this is not an inconvenience since a calibration step between the position of the marker and that of the US probe tip is required anyway in order to compute the position of the US image at runtime.

5.4.4 Robot Control

We implemented the robot control software as a C++ module for the ROS framework, indicated as "Follower" in Fig. 5.6.

The objective of the robot control routine is to update the robot position such that the ultrasound and gamma images have the maximum overlap. This is achieved when the gamma camera is looking at the portion of space immediately below the US probe.
In virtue of this, it is possible to derive the **desired robot position** ${}^{effector}_{base}T'$ as the configuration which aligns the surface of the gamma camera with the US imaging plane, the center of its upper edge to the center of the sensing surface of the ultrasound probe (what we defined in Section 5.4.2 as "*tip*"), and introduces an arbitrary chosen distance d between the two devices. We chose $d = 15cm$ after empirical evaluation of the usability of the system and sensibility to a standard amount of radiotracer for a surgical procedure. This parameter is adjustable and runtime to adapt to the current frequency of events detected by the gamma camera and to the geometric configuration of the environment. Hence, the robot position should change such that the ${}^{gamma}_{tip}T$ transform converges to the desired value ${}^{gamma}_{tip}T'$, corresponding to a translation along the z_{tip} axis of d and of $l_{gamma}/2$ along the x_{gamma} axis, where l_{gamma} is the length of the gamma camera detector.

The robot position can be specified in Cartesian space as the 6-DoF pose between the robot *base* and *flange* frames. Since the LWR iiwa is a redundant robot with 7 DoF, this specification is not univocal: multiple joint configurations can result in the same position of the last joint, while the orientation of the robot "elbow" changes. The robot firmware allows to specify this orientation along with the desired position of the end effector in Cartesian space; this can be very useful for collision avoidance or in collaborative scenarios. However, we did not take advantage of this feature in the scope of this work and only specified the target Cartesian position ${}^{effector}_{base}T'$. In this way we are able to use the advanced native Cartesian Position Controller of the robot.
In order to implement our implicit collaborative control model, we compute the target robot position ${}^{effector}_{base}T'$ according to the position of the US probe held by the operator, which is detected through the optical tracking marker *marker*. In the previous paragraph we derived

our target ${}^{gamma}_{tip}T'$ position. Since we know the ${}^{gamma}_{flange}T$ transformation by construction (as we mentioned in Section 5.4.2), we can derive the target position of the flange with respect to the current position of the ultrasound probe as

$$ {}^{flange}_{tip}T' = {}^{gamma}_{tip}T' \, {}^{flange}_{gamma}T = {}^{gamma}_{tip}T' \, {}^{gamma}_{flange}T^{-1} \tag{5.12} $$

The current position of the probe can be expressed with respect to the robot base as

$$ {}^{tip}_{base}T = {}^{flange}_{base}T \, {}^{tracking}_{flange}T \, {}^{marker}_{tracking}T \, {}^{tip}_{marker}T \tag{5.13} $$

where ${}^{flange}_{base}T$ is the current Cartesian pose of the robot as computed through its own Forward Kinematics, ${}^{tracking}_{flange}T$ is the hand-eye calibration between the robot and the optical tracking system, ${}^{marker}_{tracking}T$ is the output of the tracking system and ${}^{tip}_{marker}T$ is found via calibration, as discussed in Section 5.4.5.
Combining both equations we can express the target Cartesian position to be sent to the robot as

$$ \begin{aligned} {}^{flange}_{base}T' &= {}^{tip}_{base}T \, {}^{flange}_{tip}T' \\ &= {}^{flange}_{base}T \, {}^{tracking}_{flange}T \, {}^{marker}_{tracking}T \, {}^{tip}_{marker}T \, {}^{gamma}_{tip}T' \, {}^{gamma}_{flange}T^{-1} \end{aligned} \tag{5.14} $$

The motion target is continuously updated according to the last pose provided by the optical tracking system. The internal SmartServo control strategy provided by KUKA takes care of smoothly interpolating between the last trajectory that was being executed and the new one, resulting in a fluid motion even without adjusting the commanded positions in any way. However, if the operator holds the ultrasound probe still, the natural hand tremor and the noise on the tracking system output result in a vibration of the target position around an average center point. If the command would be transmitted to the robot without filtering, this would result in a rapid "start and stop" activation of the motors; such a situation is very harmful for the mechanical components of the robot itself. The producer of the robot uses the term "**second hand effect**" and warns about such danger. This problem can be solved by introducing an averaging of the last $n = 10$ frames received from the optical tracking system, and by introducing hysteresis: if the difference between the target position the current one is below a threshold, then the last commanded position is simply repeated indefinitely. The threshold is expressed as a limit on the difference in translation and in angle along the shortest line for the rotation. This results in the robot smoothly stopping when the US probe is held "still" by the operator, and resuming motion when its position changes significantly. This policy does not have any impact on the correctness of the image fusion, since the visualization software continues updating the image according to the last available data. Only the robot motion is suspended until necessary.

Another phenomenon that should be avoided is **instability**, which can happen in case of time delay between the sensor streams [30]. The position to be commanded to the robot is computed through a long chain of geometric transformations, combining information originating from different sensors, connected to different computers connected on a non-real-time network. For example, in an unlucky but not so uncommon case, the output of the optical tracking system would reach the control program with some lag with respect to the current

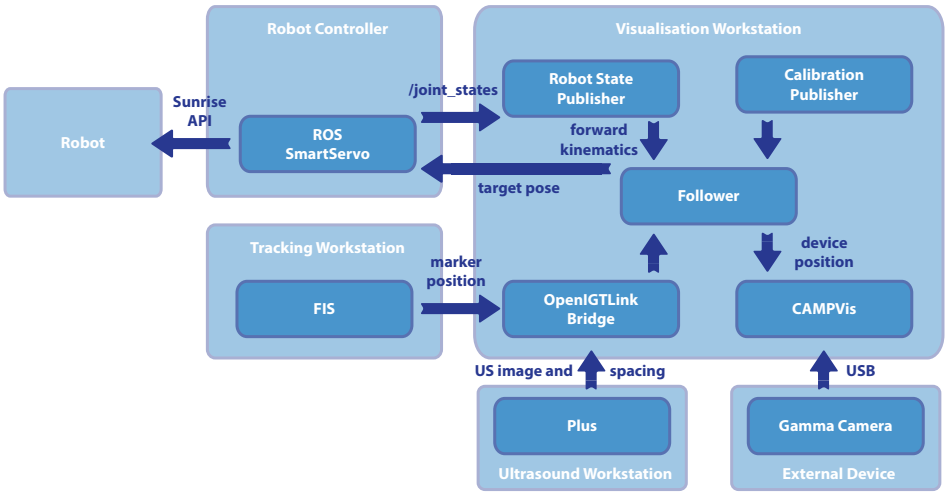


Fig. 5.6. **Architecture of the robot control system.** In this picture, FIS indicates the Optical Tracking System software.

robot position. The consequence would be an error in the commanded position that would pull the robot away from the correct target position in one direction. At the next iteration of the control routine, the new commanded position will be computed according to the updated inputs, which will be again unsynchronized. So in this unlucky case the commanded position will lead to an overcompensation in the opposite direction of the last movement. If this process repeats at a high frequency, it manifests itself as the robot starting to shake as it converges to the target configuration. Depending on the geometric configuration and the lag, the system can diverge and the shaking can reach very dangerous levels for the robot as well as for any bystander.

Since in our setup network packets must travel through a series of non-real-time hardware and software stacks, random delays can accumulate to an unbounded amount. A common countermeasure is to perform a **temporal calibration** to find the time that information takes to travel from each sensor to the location of the control routine, and then compensate for this relative lag by introducing a buffer. However, this approach can only address the average latency and not the jitter acting randomly on each single network packet. Furthermore, even the average lag may not be constant in time. This is particularly true for consumer computer networking hardware, which follows a *best effort* policy. The same can be said for consumer operating systems; soft real-time variants are available, but the effect is limited if the whole software stack is not real-time as well. Moreover, even when using only real-time hardware and software, such temporal calibration step would need to be repeated every time that the setup is changed. Even more critically, buffering the sensor information consistently across the system can be very hard as its size grows.

A simpler but effective approach is to rely on timestamps and **clock synchronization**. Software solutions capable of synchronizing the clock of different computers at the level of the operating system are freely available, the most common being NTP [122]. While most consumer operating systems are already configured to synchronize the computer's clock with standard time servers around the world, it is possible to configure them to prioritize the convergence of a few peer computers over the local network. This technique makes it possible to align multiple clocks within 1 ms of error over a LAN making use of common hardware and OSs,

and to maintain this convergence indefinitely. Once the computer clocks are synchronized, each process running on any computer can timestamp the information coming from the sensor and pass it over. The advantage of this approach is that any delay happening after assigning the timestamp will not affect the coherency of the global information when collected together; it will only introduce an overall lag corresponding to the maximum delay among all pathways. Each process can maintain a buffer of the last incoming messages in order to look for the most recent common timestamp among all channels (within a certain interval). As opposed to a hardcoded delay, this approach is more robust and flexible; the ROS framework makes wide usage of it, and in particular the tf subsystem provides utilities to adopt it to maintain a distributed representation of the geometric configuration of the system.
After noticing this resonance effect in our application, we were able to mitigate it by using NTP to synchronize the robot and tracking workstations, and the rosjava instance running on the robot as a Sunrise Application as a component of iiwa_stack.

5.4.5 Calibration

In order to ensure the accuracy of such a complex system, various calibration procedures are necessary. The tracking system must be properly calibrated in order to provide accurate positioning of the marker with respect to itself. Furthermore, the geometric transformation between the flange of the robot and the base reference frame of the tracking system must be found accurately, so that the position of the marker can be brought in the robot's base coordinate system. Finally, the relative position of the marker attached to the US probe and the image must be computed.

As a prerequisite for the two further calibrations, the **tracking system** was calibrated in terms of lens distortion, camera intrinsic parameters, and extrinsic displacement between the two cameras (also known as *stereo calibration*). The lens distortion coefficients were estimated with a four-step approach proposed by Heikkila et al [69], while the stereo camera intrinsic and extrinsic parameters were computed following the approach introduced by Zhang et al [215]. The accuracy of the tracking system has been identified to be 0.21 mm $\pm$ 0.25 mm in a previous work [21].

Next, the transformation between the robot hand and the tracking system was determined. Since the tracking system was mounted on the robot itself, this could be considered a standard *eye-on-hand* case of a **hand-eye calibration** as defined by Tsai-Lenz et al [190]. We used the existing implementation of the algorithm in the ViSP library [111]. For convenience of use, in particular for periodically recurring recalibration, we developed a ROS framework which is now publicly available (easy_handeye; see Appendix C).

Finally, the US **spatial calibration** was performed. A large number of approaches and software applications can be found in literature for determining the geometric transformation between the tracking marker attached to the US probe and the location in space of the generated image [77]. To the best of our knowledge, most of them make use of phantoms of known geometric structure which, when immersed in water, are visible in the US image as a shape which is easy to automatically segment or reliably annotate by hand. The position of the features in the image for multiple locations of the probe in space can be organized in a

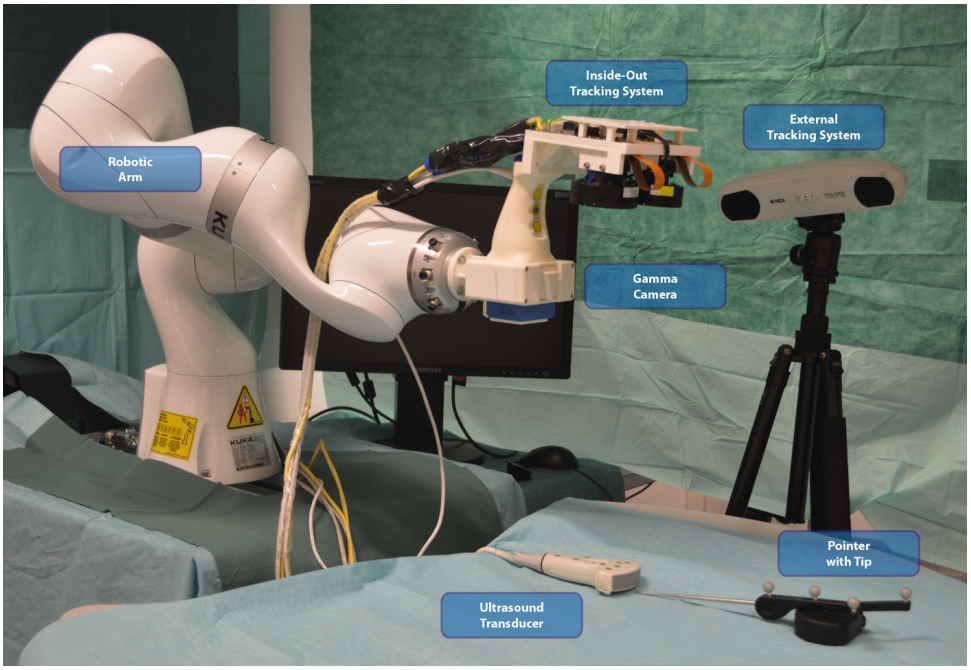


Fig. 5.7. **Ultrasound spatial calibration setup**. The system, as well as the external optical tracking system adopted to perform the calibration, are shown here. The picture is only explicative; during the actual calibration procedure, the pointer tip was touching the US probe orthogonally in the exact center of the transducer surface.

linear system of equations, which can then be solved to estimate the parameters of the transformation. However, given the noise on both the probe and feature locations, the system will not be solvable exactly. The solution is then identified as the set of parameters that leads to the minimum error. The resulting complex and error-prone procedure typically results in Point Reconstruction Accuracy no better than a millimeter [77].
In order to overcome these limitations, we adopted a simpler alternative approach. As shown in Fig. 5.7, a second, external optical tracking system (Polaris Vicra, Northern Digital Inc., Waterloo, Ontario, Canada) was employed to find the center of the transducer array of the US probe by manually imposing a pointing tool orthogonally on the probe's surface. To complete the geometric transformation we use the transformation between this point and the US image coordinate frame, provided at runtime by Plus according to the current depth setting of the machine. The whole setup is depicted in Fig. 5.7.
The spatial calibration consists in determining the transformation ${}^{image}_{marker}T$ between the tracking marker and the US image. As we mentioned, we can decompose this transformation into the relative positions of the marker and US probe tip, and between the tip and the image (defined as the top-left corner)

$$ {}^{image}_{marker}T = {}^{tip}_{marker}T \, {}^{image}_{tip}T. \tag{5.15} $$

As we mentioned, for correctly positioning the image in space ${}^{image}_{tip}T$ can be provided by the Plus OpenIGTLink server to our visualization software at runtime, once the position of the US tip is known. In order to find ${}^{tip}_{marker}T$ we use a pointing tool, consisting of a needle carrying

an optical tracking marker calibrated such that the reported pose corresponds to the tip of the needle. We can then derive that the position of the needle tip and of the US probe tip coincide:

$$ {}^{tip}_{marker}T = {}^{needle}_{marker}T. \tag{5.16} $$

In order to compute the latter transformation we need the output of the external tracking system $T^{needle}_{external}$ and the hand-eye calibration between the robot and the external tracking system $T^{external}_{base}$. The former is directly provided by the system; the latter can be found via an *eye-on-base* calibration, which we performed using easy_handeye.
Finally, we can compute the final calibration by composing the whole transformation chain:

$$ {}^{image}_{marker}T = {}^{marker}_{needle}T' = {}^{external}_{needle}T\,{}^{base}_{external}T\,{}^{flange}_{base}T\,{}^{tracking}_{flange}T\,{}^{marker}_{tracking}T, \tag{5.17} $$

where ${}^{needle}_{external}T$ is the output of the external tracking system, ${}^{external}_{base}T$ is the hand-eye calibration to the external tracking system, ${}^{flange}_{base}T$ is the pose of the robot, ${}^{tracking}_{flange}T$ is the hand-eye calibration with respect to the stereo tracking system as found before in this section, and ${}^{marker}_{tracking}T$ is the output of the stereo tracking system.
In order to minimize the impact of measurement noise, this chain of transformations was averaged over 10 samples.

It is worth noting that these calibrations remain valid until the tracking system is detached from the robot, or the marker from the US probe. The hand-eye calibration process requires 10 minutes, while the spatial US calibration requires 30 minutes. The resulting accuracy of the system was evaluated with a dedicated procedure, described in Section 5.6.

5.4.6 Image Fusion and Visualization

The visualization software is implemented as a plugin for the CAMPVis framework [169]. The plugin takes as input the current US and gamma images, their relative positions, and spatial metadata of the US image.
The grayscale US image is provided over an OpenIGTLink connection by the Plus toolkit instance running on the ultrasound machine. Since the image size and resolution is variable according to the depth setting, a matrix containing the pixel spacing of the image and the ${}^{image}_{tip}T$ translation is also received over the same connection.
A further plugin is used to directly connect to the gamma camera over USB and receive the events detected by each scintillator chamber. The events are integrated over time in order to build a grayscale image of 16 by 16 pixels, representing the event frequency for each pixel in the integration window.
A complete modeling of the gamma camera has been carried out by our group for freehand SPECT tomographic reconstruction [114]. However, for simplicity, we adopt a parallel camera model which assumes that any event detected by a "pixel" originated somewhere on the line coinciding with the axis of symmetry of the scintillator chamber.

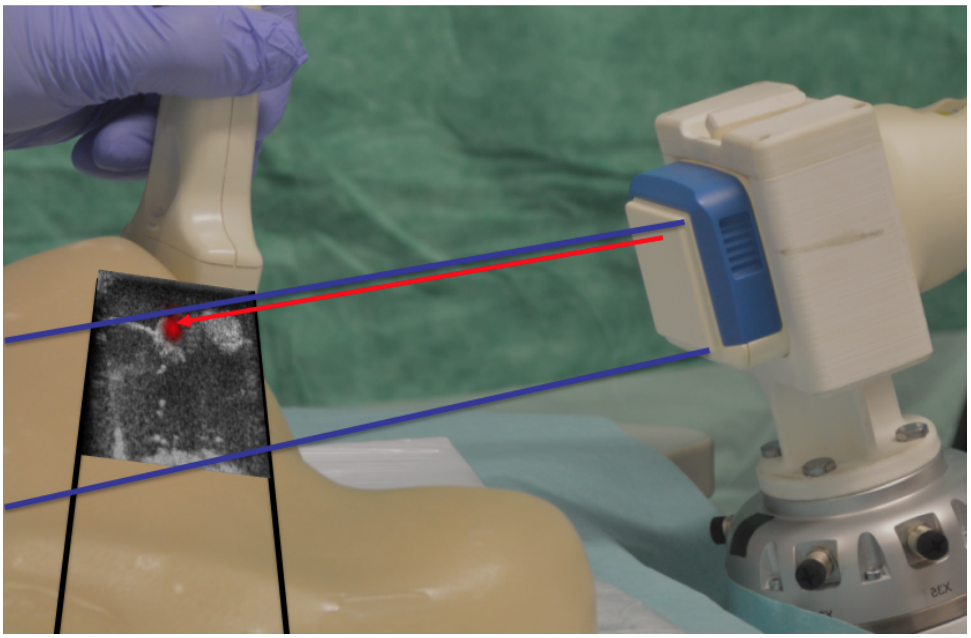


Fig. 5.8. **Visual representation of the image fusion process.** The locations of the gamma camera scintillation chambers are projected on the ultrasound image, and the number of events recorded by each of them is integrated over time. For each frame, a dot is rendered on top of the US image if the integrated event count is beyond a threshold adjustable at runtime (in order to account for the radiotracer decay).

In order to provide a familiar image to the surgeon, the US image is used as reference and the gamma camera pixels are projected on the US imaging plane using a parallel projection, as shown in Fig. 5.8. Then, the events are visualized as transparent dots, with opacity proportional to the value of the pixel in the gamma camera image. With this method, the intensity of the color is directly proportional to the frequency of the events originating from the region of space lying on a line between the gamma pixel and the corresponding US pixel. Finally, the dots are smoothened at the edges with a Gaussian profile, in order to represent the distribution of the probability in the position of the event origin and for a more interpretable and intuitive visualization.

The correct projection of the gamma events on the US image requires the knowledge of the relative position of the two images. This is computed in a separate Python script that periodically queries the tf system for the ${}^{image}_{gamma}T$ transform, and sends it to CAMPVis over an OpenIGTLink connection. In order for this to work, the whole transformation chain between the two reference frames must be known. The calibration procedure described in Section 5.4.5 provides the rigid transform between the optical marker and the tip of the ultrasound ${}^{tip}_{marker}T$. Adding the transformation ${}^{image}_{tip}T$ provided by Plus, the output of the optical tracking ${}^{marker}_{tracking}T$ and the hand-eye calibration ${}^{tracking}_{flange}T$ we can derive the transformation between the robot flange and the tip of the ultrasound probe

$$ {}^{image}_{flange}T = {}^{tracking}_{flange}T \, {}^{marker}_{tracking}T \, {}^{tip}_{marker}T \, {}^{image}_{tip}T \tag{5.18} $$

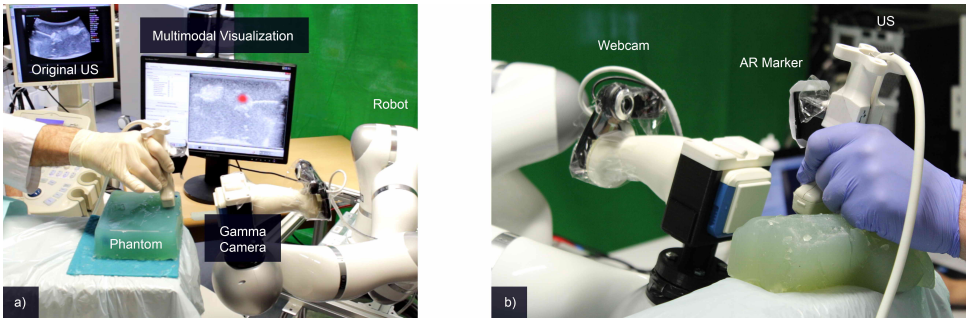


Fig. 5.9. Overview of the prototype used to test the feasibility of the concept. The first version of our system had strong hardware limitations, in particular with respect to the video tracking system. Both hardware and software were replaced completely for the second phase of this study.

As we mentioned, the position of the gamma camera scintillator with respect to the robot flange ${}^{gamma}_{flange}T$ is known by construction. Thus, we can combine this information and obtain the transformation between the two images:

$$\begin{aligned} {}^{image}_{gamma}T &= {}^{flange}_{gamma}T\,{}^{tracking}_{flange}T\,{}^{marker}_{tracking}T\,{}^{tip}_{marker}T\,{}^{image}_{tip}T \\ &= {}^{gamma}_{flange}T'\,{}^{tracking}_{flange}T\,{}^{marker}_{tracking}T\,{}^{tip}_{marker}T\,{}^{image}_{tip}T \end{aligned} \tag{5.19}$$

5.5 Feasibility Study

The first prototype of the system had a different implementation than the one described here. A simpler tracking system was used, employing a standard consumer webcam and an AR marker glued to the ultrasound probe. The robot used was also an earlier generation of the KUKA LWR, namely the LWR4+. The rest of the system, visible in Fig. 5.9, was similar in functionality.
The biggest limitation of this system was that while the robot was moving, the RGB camera image would become blurred and the tracking was lost. As a consequence, it was necessary to use a different control strategy: rather than constantly updating the commanded position, it could be computed only when the robot had reached the last target configuration. Then the robot would perform the next motion. Of course such behavior affected the performance of the system: occasionally the US probe would exit the field of view of the camera, so the robot could not see it and would stop moving until the operator would purposely bring the US probe back in sight to "guide" the robot to destination. Moreover, this piecewise movement was not an issue for small displacements of the probe, but it would hinder the usability of the system after large rotations. In this case, robot needed to rapidly go in a large circle around the probe, which led to the mentioned blurring of the RGB image and loss of tracking.

Despite these heavy limitations, the performance of the system was sufficient to test the viability of the approach. The correctness of the image fusion was tested by using a phantom consisting of a small plastic box, shown in Fig. 5.10, where spheres with the diameter of 1 cm were immersed in water. One of the two spheres contained pure water, while the other was filled with suspension of 0.5 MBq of ^{99m}Tc. The configuration of the phantom thus coarsely represented two axillary lymph nodes, of which one was labeled with the radiotracer. The aim

of the experiment was to find the minimum distance at which the SLN could be distinguished from a neighboring lymph node under our proposed multimodal imaging.
The phantom was imaged with our system, by immersing the ultrasound probe in the water in order to achieve acoustic coupling; the robot followed the probe and automatically positioned the gamma camera in front of the box, such that originating radioactivity could be detected. The two spheres were clearly visible in the US image, and one of them was clearly identifiable as the one containing the radiotracer by the red dots appearing on top in our visualization. The two dummy lymph nodes were gradually moved closer to each other in order to find the cutoff distance at which it would not have been possible to distinguish them any more; however, even when put in direct contact to each other, the source of the nuclear activity was still univocally identifiable. This fact supported the claim of a sufficient accuracy of the system for the particular application.

Furthermore, a small feasibility test with a single subject was performed on a gelatine-agar phantom including "hot" and "cold" nodes, in a setup such as the one described in the following Section 5.6. The results show the utility of the system in distinguishing the sentinel lymph node (SLN) from the non-radioactive nodes, and the ergonomy of the system while performing a needle punch biopsy (modulo the aforementioned issues). These results were reported in a paper published in MICCAI 2015 [40].

5.6 Evaluation and Validation

After a complete reengineering of the system leading to the implementation described in the rest of this chapter, a more rigorous evaluation of the accuracy of the system and its benefits in a plausible clinical scenario was executed.

5.6.1 Accuracy Evaluation

The accuracy was verified by making use of the box phantom described in the previous Section 5.5 and shown in Fig. 5.10. Again, one of the two spheres was filled with 3 MBq of ^{99m}Tc. This time however, the accuracy was statistically evaluated at different positions of the US probe with respect to the phantom in order to verify the correctness of the image fusion. Among the possible sources of error, it is worth mentioning the calibration of the tracking system (intrinsic, stereo and hand-eye), the uncertainty about the position of the gamma camera screen with respect to the robot, the spatial calibration between the ultrasound image and the tracking marker.
With the purpose of reliably positioning the US probe and holding it still, it was attached to a second robot, an UR5 (Univeral Robots, Denmark). The UR5 was operated manually through its control panel to translate and rotate the probe as much as possible while keeping the two spheres visible in the US image. After the system reached a steady state, we acquired the output multimodal image. This procedure was performed for the initial position parallel to the box wall, and then after rotating or translating in each direction for the maximum amount allowed by the geometry of the setup (i.e. without collisions with the surrounding environment). To improve the statistical significance of the experiment, we collected ten images per position.

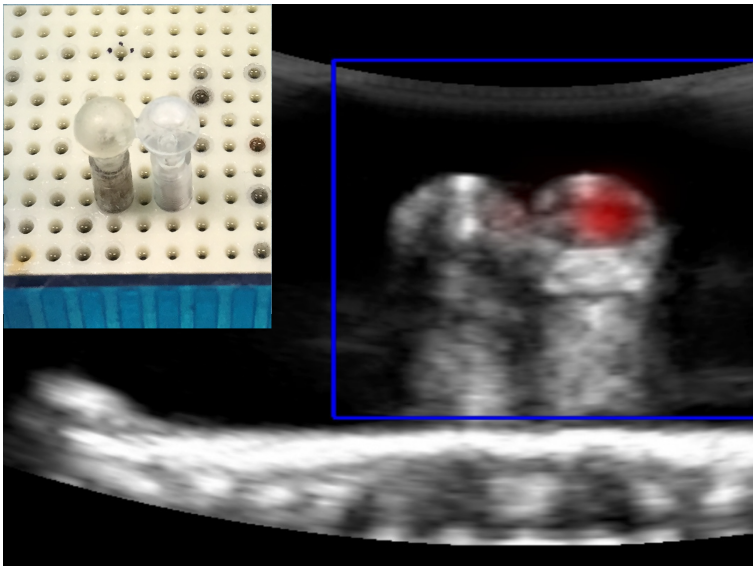


Fig. 5.10. Plastic box phantom used for evaluating the system accuracy. In the top left it is possible to see the two spheres that were immersed in water. The right one was filled with ^{99m}Tc. The rest of the figure shows the multimodal image produced by our imaging system.

On each sample image, the position of the center of the outline of the plastic sphere, visible in the US image, and of the visualized gamma activity region was manually annotated by an expert. Finally we computed the average displacement between the two points in pixels, and converted it to millimeters by measuring the apparent size in pixels of the two spheres of known size (1 cm). In order for the gamma visualization to be stable, we raised the event integration time (described in Section 5.4.6) to 2 s.
The results, reported in Table 5.1, showed that the performance of the system was reliable for any of the reached positions and orientations of the probe. Even after rotations as large as 20°, the average error on the annotated center positions was always below 1.50 mm. The median error was 1.00 mm. Considering the reported threshold of 5 mm for a LN to be visible at all in US images [140], said accuracy was deemed within the clinical requirements.

5.6.2 Expert Validation

In order to highlight the benefits of the system, a pool of five researchers and clinicians was asked to perform a needle punch biopsy under US and gamma guidance, with and without the assistance of the robot. For each procedure a protocol (reported in Tables 5.2 and 5.3) was recorded with the aim of quantitatively measuring the impact of the robot assistant on the efficiency and the outcome of the biopsy.

As for the biopsy experiment described in Section 5.5, a phantom resembling the anatomy of a human female axilla was created out of a gelatine-agar mixture as described in [33]. Some nodules (eleven in one case, twelve in another) resembling lymph nodes were immersed in the phantom before solidification. The nodules differed from the surrounding tissue for their consistency and color, and half of them were marked with the inclusion of 3 MBq of ^{99m}Tc each. While the background tissue was opaque and white (in order not to make the

Pose Number	Movement	Distance	Avg. Error
1	Initial Position	0	0.70 mm
2	Translation along y	-6 mm	1.36 mm
3	Translation along y	21 mm	0.74 mm
4	Rotation around y	$19.6°$	1.19 mm
5	Rotation around y	$17.5°$	1.39 mm
6	Rotation around x	$-22.4°$	0.93 mm
7	Rotation around x	$18.8°$	1.04 mm
8	Rotation around z	$-10.8°$	1.29 mm
9	Rotation around z	$16.3°$	1.03 mm

Tab. 5.1. **Results of the Accuracy Evaluation for varying position of the probe with respect to the phantom.** The consistency of the error for large rotation angles indicates the absence of a significant parallax error, and suggests a sufficient accuracy of the system calibration.

identification and biopsy unrealistically easy), the cold nodules were marked in blue with food coloring and the hot ones in red. A different concentration of gelatine-agar was employed in order to render a different haptic feedback between the artificial LNs and the surrounding tissue during the biopsy. While the latter consisted of a misture of 4 weight percent (wt%) of gelatine and 1.5 wt% of agar, the nodes were made out of 12 wt% and 8 wt% respectively. The lymph nodes were placed in the phantom in clusters, resembling natural dispositions that can be encountered in human anatomy, at an approximate depth of 2 cm.

	Standard Procedure				
Subject ID	1	2	3	4	5
Expertise	Professional Gynaecologist	Professional Gynaecologist	Medical Expert	Medical Expert	Medical Expert
Inspection Time (min:sec)	17:10	14:25	13:21	4:35	16:34
Nodes identified on US	11/11	10/11	9/11	9/11	9/11
Hot Nodes identified on multimodal	6	5	6	5	3
Correctly identified Hot Nodes	3/4	3/4	3/4	3/4	3/4
False positives	3	2	3	2	3
Average confidence	4.83	4	3.33	4.38	5
Biopsy Time (min:sec)	2:40	2:45	6:40	7:04	6:52
Success	no	no	no	no	no
Easiness of use	2/5	2/5	2/5	2/5	1/5
Remarks	1	1			

Tab. 5.2. **Results of the Expert Validation for the Standard Procedure without robot guidance.** The "Success" column reports if the subject managed to obtain a tissue sample of the SLN. Remarks: 1) Subjects 1 and 2 biopsied successfully a cold node situated next to the hot one that was chosen for biopsy.

	Proposed Method				
Subject ID	1	2	3	4	5
Expertise	Professional Gynaecologist	Professional Gynaecologist	Medical Expert	Medical Expert	Medical Expert
Inspection Time (min:sec)	5:46	14:42	11:23	3:53	8:52
Nodes identified on US	10/12	10/12	9/12	10/12	12/12
Hot Nodes identified on multimodal	3	5	4	5	5
Correctly identified Hot Nodes	3/5	5/5	4/5	5/5	5/5
False positives	0	0	0	0	0
Average confidence	4.67	4.6	4.13	4.2	4.4
Biopsy Time (min:sec)	1:32	2:48	4:18	2:22	2:52
Success	yes	yes	yes	no	yes
Easiness of use	4/5	5/5	4/5	5/5	4/5
Remarks				2	

Tab. 5.3. **Results of the Expert Validation for the Proposed Method with robot guidance.** Remarks: 2) After Subject 4 correctly chose a hot nodule for biopsy, while picking up the needle the US probe had a movement which led the robot against a joint limit, with consequent stop; the subject decided to biopsy the node anyway without Gamma information, and while doing so mistakingly biopsied the LN next to the chosen target.

Average		
Procedure	Standard	Proposed Method
Inspection Time (min:sec)	13:13	8:55
Nodes identified on US	9.6/11 (87.3%)	10.2/12 (85%)
Hot Nodes identified on multimodal	5	4.4
Correctly identified Hot Nodes	3/4 (75%)	4.4/5 (88%)
False positives	2.6	0
Average confidence	4.31	4.4
Biopsy Time (min:sec)	5:12	2:46
Success	0%	80%
Easiness of use	1.8	4.4

Tab. 5.4. **Average results across participants to the study.** A significant improvement is noticeable across all criteria recorded in the protocol. It should be noted that during the second phase (involving the usage of robotic imaging) two subjects spontaneously declared that in hindsight their amount of confidence expressed in locating SLNs in the first phase should have been lower.

All subjects were given time to explore the anatomy and get familiar with the system in the first place, until they felt confident to start the experiment. Then they were asked to count and identify all visible lymph nodes, and point out the radioactive ones. Their last task was to biopsy a radioactive lymph node of their own choice with a standard punch biopsy needle (HistoCore 250 mm, BIP GmbH, Germany). The success was verified by visually inspecting the color of the extracted sample.
The process was repeated twice to compare the non-assisted and assisted approaches; two different phantoms were used for the two phases, in order to avoid the use of previously acquired knowledge.
In the first round the subjects had to execute the experiment by holding the US probe and the gamma camera in their hands, while looking at two screens showing respectively the B-mode ultrasound and the gamma image. They had to mentally register the two images in order to identify the lymph node activating the gamma camera scintillator, as in the current standard medical routine. After identifying the target sLN, they had to drop the gamma camera for the US-guided biopsy to take place.
After switching phantoms, they had to repeat the procedure with robotic assistance. They could explore the anatomy by just holding the US probe while looking at the real-time multimodal image, until the SLN was identified. Then they could seamlessly perform the biopsy with the free hand, targeting the lymph node labeled in red on the grayscale anatomical image.

The experiment showed evident benefits of the multimodal imaging guidance. The full protocols are reported in Tables 5.2 and 5.3, and are summarized in Table 5.4. The participants were able to identify the hot nodes faster, with higher sensitivity and no false positives (as opposed to a high number of false positives and mislabeling without multimodal guidance). All users reported a much easier experience with the collaborative imaging system than without, and were able to use it right away. Furthermore, the success rate jumped from an astonishing zero to 80%; it is worth noting that the only failure while using the proposed system actually

happened during a system failure. After locating the hot node to be biopsied correctly, the subject got distracted while picking up the core needle and dropped the US probe. The robot tried to reach a position outside of its workspace, and in doing so it reached a joint limit and triggered an emergency stop. The subject decided to go on with the biopsy anyway, being sure of remembering which node was chosen; the color of the sample proved otherwise. This episode only strengthens the case for the importance of real-time multimodal imaging and the reduction of the mental burden on the operator through the collaborative assistance. However, it also shows that further development the prototype is necessary before the concept can be validated through clinical testing.

5.7 Outlook and Future Work

In this chapter we presented a collaborative robotic assistant, which enables intraoperative real-time multimodal imaging in the context of a workflow familiar to expert surgeons. Our experiments show the feasibility of the approach in a simulated controlled environment, as well as the strong added value provided by real-time multimodal imaging for intraoperative guidance. The accuracy reached by users within few minutes of usage of the system is impressive when compared to that of an experienced surgeon without the support of our system. This evidence indicates great potential for future applications of collaborative robotic imaging in other scenarios, as one of the few means currently known capable of providing patient-specific, real-time multimodal imaging.

However, the translation of these approach to the medical routine may be a long journey. While contemporary robots present impressive collaborative features, making them suitable to work safely next to humans, still requires extensive work on Human-Machine Interfaces in order to enable complex collaboration schemes and achieve a more complete awareness of the environment. Recent advancements in computer vision and artificial intelligence promise to overcome these challenges in the near future.
A more specific problem to the proposed approach is the need for *ad-hoc* hardware, such as a robot tailored in its geometric configuration to the particular setup. While the use of a standard industrial manipulator offered us a lot of flexibility for research, its workspace was not a perfect fit for the task at hand. A custom-made robot, maybe fixed to the wall or to the ceiling, would turn into a much better usability of the system, as it could reduce the number of instances where the arm may reach a singularity, exhaust its joint range or collide with the environment. The integration of the gamma camera and tracking system in a single encasing would also lead to a much more compact design and allow sterilization.

With the aforementioned developments, it would be interesting to perform a in-depth clinical study to verify the impact of multimodal guidance on the accuracy of core needle biopsy for breast cancer staging, and hence if it can be an enabler of its adoption as standard practice.

6 Making Ultrasound Compounding Easier and Operator-Independent through Tracking Data Denoising

In this chapter we discuss our second major contribution: an approach for improving freehand 3D ultrasound image quality by regularizing the output of the employed tracking system via a variational model, that takes into account the intrinsic geometry of the space of pose matrices. Classical ultrasound (US) compounding makes use of an external tracking system in order to co-locate US frames acquired over time with respect to a common reference frame. The values of the US frame pixels can be considered as samples of a scalar field distributed over the Cartesian space. The objective of compounding is then to regress this distribution, given the available sparse samples. This is commonly achieved by establishing a frame of reference and a a regular voxel grid covering the region of interest. For each of this voxels, the respective scalar value is determined from the samples in the neighborhood. Many different strategies to perform this reconstruction (or compounding) step have been investigated, but in general for the regression to succeed, it is very important that each pixel of the original ultrasound sweep is assigned to the correct voxel in Cartesian space. Failures to do so can yield global deformations of the image, artifacts or blur and loss of detail.

As a result, the literature about approaches addressing the improvement of compounding quality is vast. Some methods manipulate the tracking data in isolation, without exploiting additional information from external sources. Other techniques rely on image features in order to find correspondences between different US frames and improve their alignment on this basis. Other authors propose to use an additional sensor, such as an Inertial Measurement Unit (IMU), capable of augmenting the main tracking system output with complementary information about linear and angular acceleration. Approaches capable of computing the displacement between US slices from the image alone have also been presented [75, 155].

However, each of the aforementioned methodologies offers a different trade-off. Furthermore, while the majority of the contributions in the area aim at increasing the global correctness of the compounded image (typically by compensating for large-scale distortion in EM tracking), the topic of improving the perceived quality by addressing small-scale artifacts has not been as extensively addressed. We propose a solution that regularizes the tracking signal according to intrinsic geometric information, without considering US image content; but in contrast to most approaches, it includes an explicit data fidelity term. The impact of this property on ultrasound compounding is shown qualitatively and quantitatively.

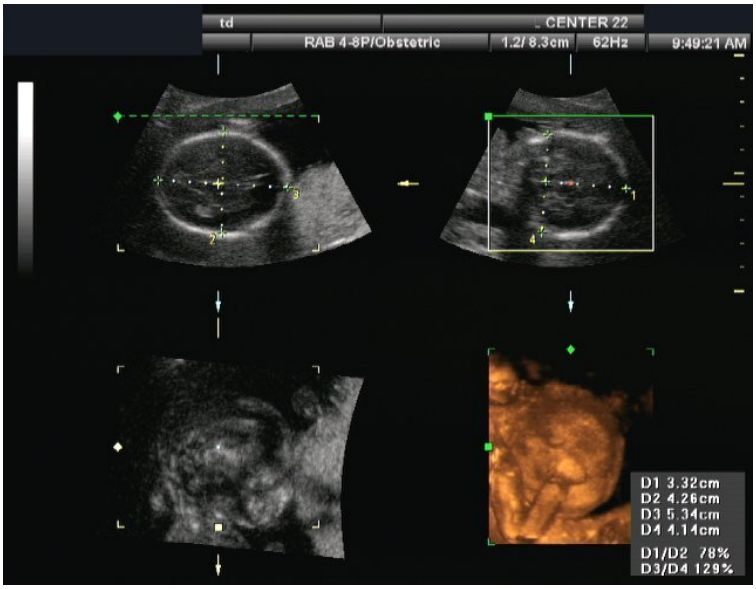


Fig. 6.1. **Measurement on 3D Ultrasound Image.** A sonographer annotates a 3D ultrasound image of a fetus' head to measure the length and width of the cranium. The 3D volume can be resliced interactively to find the plane of maximum intersection, hence reducing the chances of error with respect to 2D ultrasound, where the operator must find this volume during the acquisition itself.
Image under Creative Commons License (CC-BY-SA 3.0), by Terry J. DuBose over WikiMedia Commons.

The main contributions presented in this chapter are:

- a method for improving freehand 3D ultrasound via regularization of the pose stream provided by the tracking system, and
- a quantitative evaluation of the impact of said method on real-world data.

This work was published in IEEE Transactions on Medical Imaging (TMI) [42].

6.1 Motivation

As highlighted in the introduction to ultrasound imaging in Section 2.2.3, echography is a popular imaging modality for pre- and intraoperative usage. While surgeons employ this imaging modality in the OR because of its flexibility, speed and interactiveness, its usefulness has also been highlighted in a number of diagnostic scenarios [149]. The same characteristics make it useful in emergency cases, for example when the patient reaches the Emergency Room in a state of unconsciousness and it is vital to promptly check for potential internal bleedings or embolism. Furthermore, US imaging is preferred where ionizing radiation would bring more hazard than benefit, when a metallic implant or particular condition prevents the patient from entering an MRI device, or when it is necessary to monitor a process over time and this is not possible with another diagnostic modality.
The most frequent and widely known variant of US imaging is arguably 2D B-mode, which is mostly known to the general public because of the routine screening during pregnancy for monitoring the health and growth of the fetus. Anybody who assisted such procedure

knows how hard the resulting images are to interpret, until the physician finds the sagittal plane of the child's skull and the familiar silhouette of a baby can be recognized. During this screening, the physician usually inspects the cranium for possible conditions needing attention, such as cleft palate. Furthermore, it is common routine to measure key features, such as the axial and sagittal diameter and the circumference, for which the normal relationship has been estimated [59] and whose deviation from the known range can suggest abnormal development. Such screenings are also employed after birth in order to monitor further growth and development.

The measurement of anatomical structures via 2D ultrasound imaging is a widespread procedure, but it is also error prone [54]. The physician navigating the anatomy has to find the plane that in their estimation intersects the object of interest in the predetermined direction and along the maximum extent. The relative image must be stored and then annotated manually, so that the dimension can be derived by counting the pixels between the manually selected points and dividing them by the pixel spacing, provided by the machine according to current settings and the calibration of the probe used during the scan. With these constraints, volume can only be measured indirectly via approximated formulas for each organ (e.g. ellipsoid) involving measurements along multiple directions. The final accuracy can be undermined by a poor positioning of the US probe with respect to the organ, by an incorrect annotation of the image, and by the inherent approximation in the computation of the volume. While automated segmentation processes could alleviate the second problem, they are extremely hard to implement for echographic images because of the high average amount of speckle artifacts in the image. Nevertheless, moderate success is reported when using shape priors and algorithms developed ad-hoc for each organ or structure [139].

Recent technological progress and increase in commercial availability of native 3D ultrasound devices does not only result in better prenatal portraits, but also provides native volumetric ultrasound images. While annotating an organ or structure in a 3D image, the operator can select the location of each extremum of the segment in all dimensions concurrently (see Fig. 6.1). On the other hand, in the 2D variant the physician must acquire the image and pick the points in a later moment. Any error in the choice of the plane intersecting the structure will constrain the segment on the wrong plane, compromising the accuracy of the final measurement. Moreover, the richer spatial information provided by volumetric data may lead to more accurate automatic segmentation. If successful, the volume of interest may be directly measured rather than estimated.

Other than for measurement purposes, volumetric US acquisition offers further advantages over planar images [54]. Once a 3D volume has been acquired, it can be *resliced* to derive a 2D image at positions and orientations which could not be reached by a 2D probe, such as parallel to the skin. Alternatively, advanced 3D visualizations can be presented, in which structures and patterns become evident which could have been missed in 2D images.

A drawback of native 3D US imaging is the increased cost, due to the difficult manufacturing process. A cheaper alternative is offered by **mechanical 3D ultrasound imaging** [154], which is implemented by using US probes where a 1D transducer array (as the one used in conventional 2D ultrasound) is mounted on a rotating guide and can be made to *sweep* about its axis with a stepping motor. By collecting the 2D US frames corresponding to each position of the transducer array, it is possible to organize them in a solid volume (often in the shape of a trapezoid). Compared to the native variant, lower frame rates (typically around 1 Hz) can

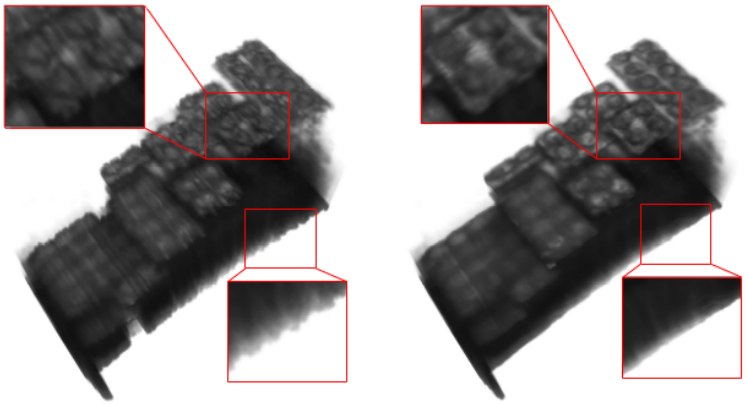


Fig. 6.2. Relationship between tracking jitter and perceived US compounding quality. The compounding of a LEGO phantom allows to appreciate the fine details of the image. It is possible to see how the regularization of the tracking data reflected on an increased smoothness of the phantom profiles. However, it is worth noting that such improvement changes local properties of the image, and does not relate to global deformations such as those caused by the interference of metallic objects with the functioning of electromagnetic tracking systems. In the latter case, the whole image is deformed in a smooth fashion. We can relate the artifacts highlighted in this image with the precision of the tracking system, while the aforementioned distortion compromises its trueness.

be achieved since a single portion of the volume can be acquired at a time. This fact can lead to motion artifacts in the image due to respiratory movements or the heartbeat.

Given the advantages offered and the usual speed of technological development in the last decades, it is legitimate to hope that in the future 3D US probes will become common hardware in clinics. However, native and mechanical 3D probes might not be sufficient to cover all possible use cases [188], since they have a limited field of view and might not be capable of capture the whole extent of a particular organ or structure. Even assuming an ever increasing upper limit to the size of the probe (due to advancing manufacturing process), a design (or number of designs) fitting to every workflow and body shape will never exist.
In this regard, **freehand 3D ultrasound** is at an advantage. Similarly to the mechanical 3D approach, it allows to combine a set of 2D frames acquired over time into a single 3D image. However, the movement is not constrained by a mechanical guide. Instead, a human operator holding the probe is required to move it freely over the region of interest. In doing so, the user can follow an arbitrary path and image an area of choice. This fact, combined with the lower cost compared to a native 3D system, makes ultrasound compounding a popular research topic, and many US machine producers have extended their devices with tracking features or have developed dedicated systems.
However, US Compounding does not come without drawbacks. Until recently, its **computational requirements** were nearly prohibitive. In time, hardware caught up and algorithms capable of fast compounding with real-time feedback have been developed [200]. Still, the quality of the final image depends in most cases on the **operator's skill**: while sweeping over the region of interest, the user must keep a constant pressure on the patient's skin, in order to avoid changes in the compression of the tissue, as this would result in artifacts in the image. Furthermore, the movement must be *regular* and consistent (some acquisition protocols even require a constant motion of the probe in the same direction).

Among all possible error factors in freehand 3D ultrasound, we can distinguish defects in the US image acquisition and digitalization, temporal matching between probe tracking data and the corresponding image, spatial calibration between the marker and the image, and reconstruction errors [188]. While such issues can be addressed individually with better algorithmic solutions or *ad-hoc* procedures, there is one matter which requires to evaluate a trade-off between the compounding quality and the usability of the system: the choice of the technology used for **tracking** the probe.

Early prototypes used **mechanical tracking**: a passive robotic arm was attached to the probe to record its position in space [55]. While mechanical tracking can achieve high frequency and precision, its usability is not optimal: the robotic arm can be bulky, heavy to move, and the movement necessary to perform the scan could lead out of its dexterous workspace (as a result, it would block and the physician would have to move the patient and start from scratch). These drawbacks can result in a frustrating experience and a heavy disruption of the clinical workflow.

A "wireless" alternative is offered by **optical tracking**, which employs video cameras to track the position of visual markers attached the US probe [189]. The markers can be reflective objects of known geometry, or can emit light of a predetermined color to facilitate image segmentation (this can be achieved by using LEDs); the former case is often defined *passive tracking* in contrast to the latter, called *active tracking*, which requires an extra cable connecting the marker to the tracking system. While there is no mechanical mean impeding the movement of the probe, the tracking system has a limited workspace (corresponding to the field of view of the camera where the size of the markers allows their detection) and, most notably, a *line-of-sight* requirement: the marker must be visible to the camera at all times, or the tracking will be lost. Depending on the compounding algorithm, this could result in image artifacts or in an irrecoverable failure. As a consequence the operator must pay constant attention not to interpose their person between the camera and the marker; or, depending on its geometry, not to rotate the probe in a way to hide the marker from the tracking system. This is, of course, a major annoyance and distraction from the task at hand.

On the other hand, **electromagnetic tracking** is capable of good accuracy without line-of-sight constraints [79]. It requires to position the field generator in the proximity of the operating site and to attach a sensor to the probe (some producers offer US probes with an integrated and pre-calibrated EM sensor), but the movement of the latter is otherwise not constrained in any physical way. The workspace of the EM system is limited, but it is sufficient for most use cases. The main disadvantage of EM tracking is the sensibility to magnetic or metallic objects in its proximity, which can lead to a degradation of the tracking accuracy. Depending on the scenario, systematic error can be introduced as well as non-systematic. Preventing such interference requires a metallic-free environment; such requirement can be challenging to fulfill, since most hospital beds include a metallic frame. However, the better usability makes EM tracking the preferred alternative for inclusion in most commercial systems as of today.

Another notable trend is **sensorless freehand 3D ultrasound**, which derives the relative displacement between US frames acquired consecutively from local image features or statistics. Many existing approaches exploit speckle decorrelation [129], and while the first methods required to perform the sweep in a constant direction because of the ambiguity in the computed movement direction, this limitation has been overcome by recent developments [75]. Recently, learning-based methods have shown promising results [155]. However, an intrinsic limitation of said approaches is their incremental nature. Since the probe displacement is computed between pair of consecutive images, the error accumulating over the image series can lead

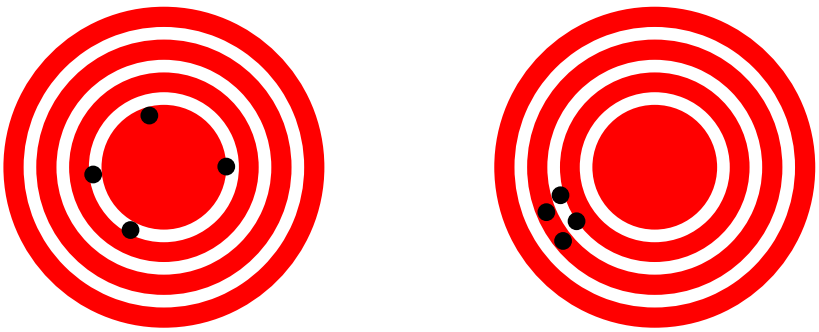


Fig. 6.3. **Trueness and Precision as components of Accuracy.** According to the ISO 5725 Standard of 1994, the quality of a measurement can be globally defined as *accuracy*. However, the accuracy of a system can only be quantified by comparing a set of measurements against a known *ground truth*, and two distinct features of a set of measurements can be identified which will impact the overall accuracy. *Trueness* reflects how close the set of measurements is on average with respect to the ground truth. *Precision* summarizes how close the measurements are to each other. This figure helps explain this dicotomy: the left panel shows four measurements which are not close to each other, but whose center of mass lies close to the ground truth. On the other hand, the set of sample points in the right half is more compact, but the average coordinates lie further away from the ground truth than the previous example. If these were representing the output of two systems, the first would be characterized by high trueness, but low precision; viceversa for the second one. It is, however, not possible to conclude that one is more *accurate* than the other without a quantitative analysis.
Original image by DarkEvil (WikiMedia Commons) under Public Domain.

to a significant overall positional error at a distance. While this fact may impact negatively the accuracy of the measurement of anatomical structures, these methods appear to be less susceptible to jitter and are potentially capable of providing more pleasant images which could be suitable for qualitative inspection by the physician [75]. In the scope of this work, we will focus on tracking-based approaches which are more robust to drift but are influenced by jitter.

As with any technology, tracking systems are not perfect and are subject to different kinds of error. The error might be systematic, such as a **distortion** [51] of the reported coordinates of the object across the workspace. A classic example is the interference of a big metallic object in the workspace of an EM system, which may "bend" the output trajectory of an object moved in a straight trajectory in space. Such error is more subtle than jitter, as it can not be detected without the aid of an external device (such as another tracking system immune to the interference, or a way to move an object in a straight line), but it can lead to grossly inaccurate measurements on the anatomy depicted in the (distorted) compounded volume. In virtue of this, its compensation requires careful calibration of each manufactured component, shielding, or an empirical modeling of the distortion for the individual surgical setup.
Alternatively, the reported position of the marker may disagree with the truth in a way which is neither predictable nor homogeneous across space, and hence cannot be modeled and compensated with a simple calibration. The performance of the tracking system in this regard may be considered as a result of the intrinsic quality of the technology adopted and of the device manufacturing.
Finally, all of the aforementioned tracking technologies are subject to some degree of **jitter**:

the output for the same position of the tracked object in space will not remain constant in time, but will be subject to *noise* and vary around a center position.
The performance of a tracking system can be measured by using a physical guide or reference object to move the marker in space and compare the position as measured by the system with the ground truth. At each position, multiple samples of the tracking output should be acquired in order to be able to decompose the total error into **systematic error** and **random error**. The former reflects distortions in the environment and other error intrinsic to the technology or device, while the latter mostly manifests itself as jitter. The measurement of the systematic error evaluates the *trueness* of the system, while its *precision* reflects the noise level. Aggregated, they both limit the overall *accuracy* of the tracking device.
While all the above sources of error contribute to the final result of ultrasound compounding, jitter plays a particular role for the perceived image *quality* [95, 209]. As a result, we will focus on improving tracking *precision*, in contrast to distortion compensation methods which address mostly *trueness*.

Since most of the literature focuses on distortion correction in the context of EM tracking [51, 209], also the studies of its application to ultrasound compounding usually evaluate the global positioning error of features identifiable in the compounded volume. For example, ad-hoc phantoms including straight lines or sharp corners are imaged, so that notable points can be automatically segmented or manually annotated. The measurement result is then compared to the ground truth.
Such approaches reflect in particular the system calibration and the *accuracy* of the employed tracking system. By contrast, the tracking *precision* does not affect the global positioning of US images in the volume of interest (and hence of the fiducials used to evaluate the accuracy as above), but rather their reciprocal co-location. If the tracking precision is low, the result will hence be a low-quality image, hard to read although globally "correct" [95, 209]. This effect is noticeable in Fig. 6.2.
It is also worth noting how the perceived quality of US Compounding is affected by both the translational and rotational component of tracking noise, since the center of rotation is the position of the tracking marker which is typically located at a considerable distance from the center of the US image. Assuming such distance to be 10 cm, a rotational perturbation of $1°$ (typical for many commercial systems [51]) would lead to a translational displacement of 1.7 mm (much worse than the usual accuracy for translation).
As of today, EM tracking is reported to offer inferior performance compared to optical technologies in terms of both trueness and precision [14, 95, 209]. While preventing systematic error requires interventions on the tracking device itself and on the operating site, reduction of jitter may be achieved through post-processing of the tracking data [51].

In this Chapter, we will focus on improving the local quality of ultrasound compounding via regularization of the tracking pose stream. In the Section 6.2 we will briefly discuss the existing regularization approaches, most of them relying on hybrid tracking and data fusion (with additional hardware complexity), or on Extended Kalman Filtering (with associated "smoothing" of the object trajectory). In Section 6.4 we will expose our proposed method for processing the tracking system output, capable of reducing noise while preserving sharp direction changes in the object's path. Finally, we will describe our experiments aimed at measuring its impact on the local quality of compounded ultrasound volumes, with comparison between mechanical, optical and electromagnetic tracking.

6.2 State of the Art

As mentioned in the previous section, most articles about error correction for EM tracking systems focuses on **distortion compensation**. An approach is to perform a static calibration with respect to another tracking system considered as gold standard; success has been reported in the adoption of 2nd/3rd degree polynomial models [209]. However, such an approach is not robust to dynamic environments where the position of metallic objects and their interference can change over time; so there is strong interest in a continuous identification of distortion in order to enable real-time correction. Approaches have been proposed which fuse EM data with additional inertial or optical [192] tracking systems, or redundant EM sensors [161].
Distortion correction is not in the scope of this work, as we rather focus on minimization of tracking noise; it can be theorized that adopting one such approach jointly with our proposed method would have a complementary impact on the compounding accuracy at a global and local level.

Most efforts aiming at reducing tracking **jitter** focus on the translational component, and address the issue with **smoothing filters** [209]. Moving average, logistic regression and Extended Kalman Filter have been reported to increase the average tracking accuracy and to have a positive effect on the precision during to their smoothing effect; however, a **trade-off** between the output regularity and the reactivity to abrupt direction changes exists [209]. On the other hand, Kalman Filter-based approaches offer an opportunity to exploit prior knowledge about the noise model; cases where the motion is constrained, such as in EM-tracked intrasvascular catheters, can benefit greatly from non-holonomic formulations of the noise process [106]. However, Kalman Filter methods operate only in one direction, deriving the result only from the input data; usual formulations do not constrain the output to establish some property, such as regularity.
A particularly elegant solution to the problem can be found in the fusion of data provided by EM tracking and image-based tracking (i.e. the "sensorless" methods we mentioned in the previous section). Such methods combine perfectly the complementary strengths and weaknesses of each technology: EM is more robust to drift but subject to jitter, while the opposite is true for image-based tracking [95]. The only drawback of such approach can be found in the sensitivity to error in both spatial and temporal calibration (which can compromise the agreement between the two tracking streams).

In contrast to existing approaches, the proposed method allows to reduce tracking jitter while relying on a single source of truth (hence removing the dependency on spatial and temporal calibration between tracking and US image) and penalizing any excessive smoothing of the output trajectory via an explicit data fidelity term. In the next Sections, we will expose the mathematical foundation, the application to ultrasound compounding and the impact on final image quality.

6.3 Technologies

In the following sections we will compare the image quality obtainable with different tracking systems. Here we will introduce the tracking technologies and compounding method that we adopted.

6.3.1 Tracking systems

Given its characteristics, our method can leverage its advantages best when applied to the output of **electromagnetic tracking**. For this kind of tracking system, a *field generator* composed of multiple coils is employed to generate a magnetic field in the surroundings. The *sensors* typically also include inductors, which can convert variations in the intensity of the magnetic field into an electrical current. The current can then be measured and employed to quantify the position of the sensor in one degree of freedom (DoF) in space. It follows that the inclusion of multiple coils in the sensor and/or periodic variation of the magnetic field orientation and intensity encoding spatial information are necessary in order to implement translational and rotational 6-DoF tracking. Different technologies have been developed over time, with different strengths and weaknesses [51].
A major design decision involves the number of coils included in the sensor [191]. Multiple coils of different shape and size allow to employ different frequencies to encode spatial information related different directions and sample them concurrently; this scheme is called Frequency Division Multiplexing (FDM). On the other hand, it is possible to use a single pair of coils, and to change the magnetic field periodically to encode spatial information in different directions. Time Division Multiplexing (TDM) leads to a smaller size of the sensor, but comes at a cost: the field generator and sensor coils take time to adjust from one phase to the others, which significantly reduces the obtainable sampling rates. Moreover, since the measurements relative to the different directions take place consecutively, if the sensor is moving quickly these measurements may not refer to the same position and speed of the sensor; this fact can greatly impact the tracking accuracy.
The design of the field generator can also vary greatly, in particular to adapt to specific surgical workspaces. For example, a flat field generator can fit between the patient and the bed, and its robustness to the distortion provoked by the latter can be optimized; or it can be integrated directly into the bed itself.

We already introduced **stereo optical tracking** in Section 5.4.3. In this context we did not use a custom system, but rather a commercial device making use of spherical markers. However, the working principle is similar and the same considerations still hold.

We also used the same industrial manipulator presented in Chapter 5 to implement **mechanical tracking**. The position was computed using Forward Kinematics, as we discussed in Section 5.3.4.

6.3.2 3D Ultrasound Compounding

The freehand 3D ultrasound compounding process consists in the reconstruction of a volumetric image from a (usually sparse) set of 2D samples, whose positions and orientations in space are known.
Most approaches start by determining the position and size of the voxel grid, which will be filled with grayscale values that are computed from the samples associated to the respective voxel. This step can already influence the final image quality, and it has been object of extensive research. A common method is to consider the bounding box of the samples, and to use PCA to find a coordinate system which aligns optimally with the US slices, in order to minimize aliasing effects [162].
The next step is to compute the intensity value for each voxel. The plethora of available algorithms can be classified according to the adopted strategy [129]. One common criterion [5] distinguishes between methods scanning the 3D grid and looking up the nearest US slices (Voxel-Based Methods), others going over the 2D slice pixels to fill the volume voxels (Pixel-Based Methods), and a final category of algorithms which use the 2D pixel values to fit 3D functions, and then sample them in correspondence of the grid voxels (Function-Based Methods).
The most straightforward VBM is a naive nearest-neighbor strategy (Voxel Nearest-Neighbor). The resulting crisp artifacts can be attenuated by adopting an interpolation scheme to average out among pixels in the neighborhood.
PBMs can vary between naive value assignment based on distance, to adopting sophisticated, physically grounded propagation models. It is common to follow the value assignment with a further scan of the volume to close gaps due to the sparsity of the samples.
The kernel functions that have been proposed for adoption in FBM approaches include Radial Basis Functions and Bayesian frameworks.
Some authors propose schemes that do not only exploit geometric information and prior knowledge about the underlying physical phenomenon, but also consider the content of the 2D US slices to improve image quality and reduce artifacts.
More detail about the individual methods can be found in comprehensive review papers [119, 129, 175].

6.4 Proposed Method

Here we propose a method for the regularization of the tracking data (such as the synthetic example in Fig. 6.4), with the aim of reducing jitter in order to improve the reciprocal co-location of the 2D slice in space and in turn improve the perceived quality of the volumetric image.
The approach consists in a pre-processing step, that takes the original tracking data as input and provides a regularized version as its output. While it could be advantageous to first apply a distortion correction approach (as mentioned in Section 6.1), this is outside the scope of the presented work.
The method is intended to be general, and to improve the quality of the compounding obtained with any reconstruction method among those mentioned in the previous section. We will perform our evaluation using a straightforward compounding approach (averaging of pixels

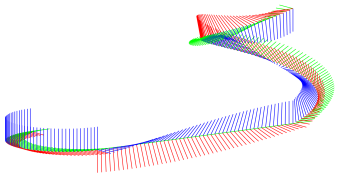


Fig. 6.4. **Ground Truth Signal** used for synthetic experiments. All poses are shown as colored tripods, where each colored leg indicates the position and orientation of the respective unit vector that has been transformed by the corresponding pose matrix $\mathbf{p} \in SE(3)$: red - x axis, green - y axis, blue - z axis.

within a distance threshold), but of course the quality may be further improved by adopting a more advanced compounding algorithm.

Before we introduce the model and discuss its derivation, we need to define our notation and discuss some necessary theoretical concepts. In order to prevent the presentation from becoming too technical, we defer the mathematical concepts, which are necessary for the implementation, to Appendix D.

6.4.1 The Special Euclidean Group

The output of a tracking system is a series of poses or Euclidean transformations consisting of a rotation matrix R and a translation vector t, c.f. also Fig. 6.4. The set of all tuples (R, t) forms the special Euclidean group $SE(3)$ which is a semi-direct product of the rotation group $SO(3)$ and the translation group. This means that group operations involve one of the rotational components acting on the translational component of the other tuple when carrying out the group operation. This becomes evident when representing $SE(3)$ in the following form:

$$SE(3) = \left\{ \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix} : R \in SO(3), t \in \mathbb{R}^3 \right\}. \tag{6.1}$$

Thus, $SE(3)$ is a subgroup of the group of invertible 4×4 matrices $GL(4, \mathbb{R})$ with matrix multiplication as canonical group operation. We emphasize this important point since it has immediate consequences on the computations to be carried out.

We consider the manifold $\mathcal{M} = SE(3)$ of Euclidean transformations endowed with the Riemannian metric discussed in Section 6.4.2. For a pose $\mathbf{p} \in \mathcal{M}$ we denote the tangent space at $\mathbf{p}$ by $T_{\mathbf{p}}\mathcal{M}$. Then we consider the exponential mapping

$$\exp_{\mathbf{p}} : \begin{cases} T_{\mathbf{p}}\mathcal{M} & \to \mathcal{M}, \\ v & \mapsto \exp_{\mathbf{p}}(v), \end{cases} \tag{6.2}$$

where $v \in T_{\mathbf{p}}\mathcal{M}$, which maps at tangent vector at v in the tangent space of the pose $\mathbf{p}$ to a new pose $\exp_{\mathbf{p}}(v)$ by following the geodesic starting at $\mathbf{p}$ with direction v. For the implementation

of this operation, we refer the reader to the appendix. The inverse exponential mapping at $\mathbf{p}$

$$\log_{\mathbf{p}} : \begin{cases} \mathcal{M} & \to T_{\mathbf{p}}\mathcal{M}, \\ \mathbf{q} & \mapsto \log_{\mathbf{p}}(\mathbf{q}), \end{cases} \tag{6.3}$$

maps a pose $\mathbf{q}$ to the tangent vector $\log_{\mathbf{p}}(\mathbf{q})$ in $T_{\mathbf{p}}\mathcal{M}$ (infinitesimal rotation and translation) located at $\mathbf{p}$ and pointing towards the pose $\mathbf{q}$. Then, we have that

$$\exp_{\mathbf{p}}(\log_{\mathbf{p}}(\mathbf{q})) = \mathbf{q}. \tag{6.4}$$

A shortest path on $\mathcal{M}$ between two poses $\mathbf{p}$ and $\mathbf{q}$ is called a geodesic and the distance between $\mathbf{p}$ and $\mathbf{q}$ is given by length of this geodesic. Hence, if the inverse exponential mapping is unique, we have for the distance $d(\mathbf{p}, \mathbf{q})$ between $\mathbf{p}$ and $\mathbf{q}$,

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{\mathfrak{g}(\log_{\mathbf{p}}(\mathbf{q}), \log_{\mathbf{p}}(\mathbf{q}))}, \tag{6.5}$$

where

$$\mathfrak{g} : \begin{cases} T_{\mathbf{p}}\mathcal{M} \times T_{\mathbf{p}}\mathcal{M} & \to [0, \infty), \\ (v, w) & \mapsto \mathfrak{g}(v, w) \end{cases} \tag{6.6}$$

denotes the chosen Riemannian metric $\mathfrak{g}$.

6.4.2 Choice of the Riemannian Metric

In order to define a Riemannian metric, we have to take a closer look at the elements of $T_{\mathbf{p}}\mathcal{M}$: any $v \in T_{\mathbf{p}}\mathcal{M}$ takes the form

$$v = \begin{pmatrix} \omega_v & t_v \\ 0 & 0 \end{pmatrix}, \tag{6.7}$$

where $t_v \in \mathbb{R}^3$ and

$$\omega_v = \begin{pmatrix} 0 & -\omega_v^z & \omega_v^y \\ \omega_v^z & 0 & -\omega_v^x \\ -\omega_v^y & \omega_v^x & 0 \end{pmatrix}. \tag{6.8}$$

Here $\omega_v^x, \omega_y^v, \omega_z^v$ denote the infinitesimal angular displacements with respect to the corresponding axis. An appropriate Riemannian metric for $v, w \in T_{\mathbf{p}}\mathcal{M}$ is given by

$$\mathfrak{g}_{\mathbf{p}}(v, w) = \text{trace}(\omega_v^T \omega_w) + t_v \cdot t_w. \tag{6.9}$$

As a consequence, we find that for a pose

$$\mathbf{q} = \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix}, \tag{6.10}$$

we have

$$\mathfrak{g}_{\mathbf{qp}}(\mathbf{q}v, \mathbf{q}w) = \text{trace}(\omega_v^T R^T R \omega_w) + t_v^T R^T R t_w = \mathfrak{g}_{\mathbf{p}}(v, w). \tag{6.11}$$

Thus, Eq. 6.9 is a left-invariant metric, i.e., it is invariant with respect to the choice of a global coordinate frame. However, $\mathfrak{g}$ is not right-invariant, which means that it is not invariant regarding the choice of a body-fixed reference frame. In the context of 3D freehand ultrasound compounding, this means that by using this metric the computed results are independent of the choice of a global reference frame, such as the one provided by the tracking system itself, but not independent regarding the calibration of the ultrasound transducer. It is important to understand that the proposed methodology can also be combined with right-invariant metrics if desired. Achieving both left- and right-invariance is, however, not possible, but this is not a shortcoming of the proposed method, rather than an intrinsic structural issue of $SE(3)$ (which is neither compact nor commutative). In this work, we focus on the metric defined in Eq. 6.9 and refer the interested reader to [8, 20, 148, 210, 211] for further reading.

6.4.3 Regularization Model for Pose Signals

In this section we describe a general framework for regularizing pose signals, which is conceptually similar to the work of Weinmann *et al.* [199]. To this end, we consider a pose(-valued) signal with $k \in \mathbb{N}$ entries as a vector $\mathbf{p} = (\mathbf{p}_1, \dots, \mathbf{p}_k) \in \mathcal{M}^k$. We assume that $\mathbf{p}$ is given, e.g., obtained by an optical or EM tracking system, and we wish to find another signal $\mathbf{x}$ such that the functional

$$E(\mathbf{x}) = D(\mathbf{x}, \mathbf{p}) + \alpha R(\mathbf{x}), \quad \alpha > 0, \tag{6.12}$$

is minimal. While $D(\mathbf{x}, \mathbf{p})$ is a data fidelity term penalizing the deviation from the original signal $\mathbf{p}$, $R(\mathbf{x})$ is a regularizer penalizing large variations or jumps in $\mathbf{x}$. We also investigated the impact of a second-order regularizer $R_2(\mathbf{x})$, which has been considered by Bacák *et al.* [2]. The associated regularization parameter is denoted by β and its influence is investigated briefly in the experiments, cf. Fig. 6.5. As its impact did not significantly change the results, we deferred its discussion to the appendix D.1.

6.4.4 Data Term

As a data fidelity term, we consider

$$D(\mathbf{x}, \mathbf{f}) = \sum_{i=1}^{k} (h \circ d)(\mathbf{x}_i, \mathbf{p}_i), \tag{6.13}$$

where d is the metric defined in Eq. 6.5 and h is one of the following functions: $h(s) = s$ which leads to an ℓ_1-type penalization, $h(s) = s^2/2$ which leads to an ℓ_2-type penalization, and

$$h(s) = \begin{cases} s^2, & s < 1/\sqrt{2}, \\ \sqrt{2}s - 1/2, & \text{otherwise}, \end{cases} \tag{6.14}$$

which yields the manifold-valued equivalent of the well-known Huber-norm [78] – a differentiable compromise between the ℓ_1-norm and the ℓ_2-norm.

Algorithm 1: Cyclic proximal point algorithm for solving the functional Eq. 6.12.

input : Signal $\mathbf{p}$, parameters α and β, number of steps l, and chosen weighting function h
output : Signal $\mathbf{x}$ (solution of Eq. 6.12)

$\mathbf{x} \leftarrow \mathbf{p}$
for $j \leftarrow 1$ **to** l **do**
  //compute relaxation parameter
  $\lambda_j \leftarrow$ compLambda(j);
  //proximal mapping of D
  for $i \leftarrow 1$ **to** k **do**
    $\mathbf{x}_i \leftarrow$ proxData(λ_j,$\mathbf{x}_i$,$\mathbf{p}_i$);
  end
  //proximal mapping of R_1
  for $i \leftarrow 1$ **to** $k-1$ **do**
    $\mathbf{x}_i$,$\mathbf{x}_{i+1} \leftarrow$ proxR1(λ_j,α,$\mathbf{x}_i$,$\mathbf{x}_{i+1}$)
  end
  //if used: proximal mapping of R_2
  for $i \leftarrow 2$ **to** $k-1$ **do**
    $\mathbf{x}_{i-1}$,$\mathbf{x}_i$,$\mathbf{x}_{i+1} \leftarrow$ proxR2(λ_j,β,$\mathbf{x}_{i-1}$,$\mathbf{x}_i$,$\mathbf{x}_{i+1}$)
  end
end

In order to distinguish these three cases in an intuitive manner, we denote the case $h(s) = s$ by ℓ_1, the case $h(s) = s^2/2$ by ℓ_2 and the case of Eq. 6.14 by HUBER.

6.4.5 First Order Total Variation Regularizer

Similar to D we consider the regularizer

$$R(\mathbf{x}) = \sum_{i=1}^{k-1} (h \circ d)(\mathbf{x}_i, \mathbf{x}_{i+1}). \tag{6.15}$$

As $d(\mathbf{x}_i, \mathbf{x}_{i+1})$ can be considered as (the weighted magnitude of) a manifold-valued, first-order forward difference, R can be interpreted as a first order approximation of the classical Tikhonov regularizer, in case of $h(s) = s^2/2$, or the total variation, in case of $h(s) = s$, respectively. In case of Eq. 6.14, R can be regarded as a pose-valued differentiable approximation of the total variation regularizer, which can be used to avoid the staircasing problem associated with total variation denoising, cf. Chambolle and Pock [152]. Similar to the data term, we use abbreviations for the different regularization types, i.e. TV ($\ell_1 - case$), TKHV ($\ell_2 - case$), and HUBER.

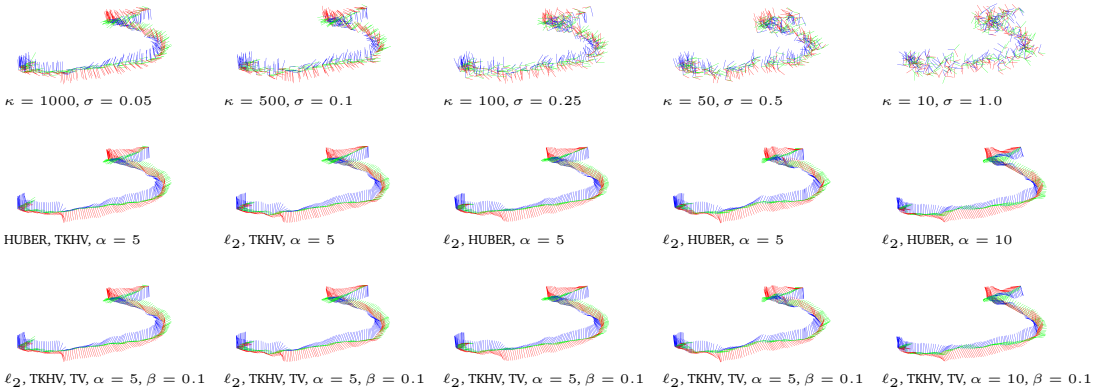


Fig. 6.5. **Visual Comparison of 1st and 2nd Order Regularization Results.** Synthetic data with various noise levels as described in Section 6.4.7 (first row), best results obtained with only 1st order regularization (second row), best results with added 2nd order regularization (third row). Optimal parameters obtained via grid search are given below the individual images, where the reading order is: data term regularizer, first order regularizer, second order regularizer and respective regularization parameters. We observe that second order regularization does not yield a significant improvement.

6.4.6 Numerical Solution

Minimizing the functional in Eq. 6.12 can be achieved by a cyclic proximal point algorithm. This algorithm consists of a cyclic application of proximal mappings to the individual atomic data and regularization terms. For a detailed derivation as well as an analysis of this algorithm we refer the reader to [199] and [2]. A pseudo-code is given in Algorithm 1. The function `compLambda` determines the optimal step size and the functions `proxData` add `proxR1` realize the proximal mappings on $SE(3)$ for Eq. 6.13 and Eq. 6.15. To keep the presentation concise, we defer all implementation details regarding these functions to the appendix.

6.4.7 Parameter Selection

The aforementioned Algorithm 1 requires the input of two parameters, α and β. We performed a grid search to find the parameters leading to the best minimization of the noise artificially introduced on a synthetic dataset. We added noise to the translational as well as the rotational components of each pose of the trajectory depicted in Fig. 6.4, as shown in the first row of Fig. 6.5. While adding noise to the translational component can be done by adding white Gaussian noise, adding noise to the rotational component is slightly more involved. To this end, we employed Gibbs sampling for a vector-valued von Mises-Fisher distribution as proposed by [16]. The respective probability density function is proportional to

$$\exp((\kappa, 0, 0, 0)^T \cdot p) \text{ for } \|p\|_2 = 1, \tag{6.16}$$

where $p \in \mathbf{S}^3$ denotes the directional component of a pose $\mathbf{p}$. By choosing $\kappa = 1000, 500, 100, 50, 10$ and $\sigma = 0.05, 0.1, 0.25, 0.5, 1.0$ (for the Gaussian noise on the translational part) we obtain the five test data sets depicted in the first row of Fig. 6.5. For evaluating the regularization performance of the proposed algorithm, we use the mean error between two

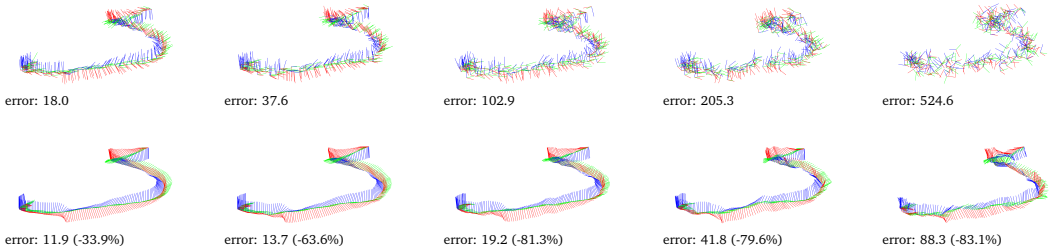


Fig. 6.6. **Results Obtained with Determined Parameters.** Synthetic data with various noise levels as also shown in Fig. 6.5 (first row) and results obtained with 1st order regularization and determined parameter settings of $p = 2$, $q = 0$ and $\alpha = 5$ (second row). Errors to ground truth signal and corresponding reduction rate are reported w.r.t. square root of the average geodesic error to the ground truth signal computed over the whole signal.

pose signals $\mathbf{p} = (\mathbf{p}_1, \ldots, \mathbf{p}_k)$ and $\mathbf{q} = (\mathbf{q}_1, \ldots, \mathbf{q}_k)$ induced by the distance metric defined in Eq. 6.5:

$$\frac{1}{k} \sum_{i=1}^{k} d(\mathbf{p}_i, \mathbf{q}_i). \tag{6.17}$$

In order to determine meaningful parameter settings, we performed a grid search with the various weightings for the data term, i.e. ℓ_1, ℓ_2 and HUBER, different regularization types, i.e. TV, TKHV and HUBER, as well as various choices for $\alpha = 0.01, 0.1, 0.5, 1, 2, 5, 10, 20$ and $\beta = 0, 0.1, 0.5, 1, 2, 5$, where we set the number of steps to 1000. Based on these experiments, we learned that the second order regularization requires smaller regularization parameter values and does not lead to significant improvements in terms of regularization performance (compare second and third row of Fig. 6.5). Thus, we concluded that an ℓ_2-type data term in combination with a first order Huber-TV-regularization seems to be a reasonable choice which performs well over a broad range of noise levels and performed another set of experiments with these settings shown in Fig. 6.6. These recommended settings lead to a significant reduction of the error between the ground truth signal and the regularized ones as demonstrated by the reported errors and reduction rates in Fig. 6.6.

6.4.8 Application to freehand 3D Ultrasound

Through the experiments described in Section 6.4.7 we determined a set of parameters that were suitable for a consistent reduction of noise that was artificially induced on a synthetic trajectory. We used the same method and parameters to regularize the pose streams provided by the tracking systems that we adopted in our custom freehand ultrasound platform, described in Section 6.5.1.

In the next Section 6.5, we will verify if the parameters found through a grid search on a synthetic dataset also translate into improvements for real world data. For this purpose we will statistically analyze and compare the output of different tracking systems before and after regularization. Furthermore, we will qualitatively and quantitatively verify the improvement of the image quality provided by US compounding reconstructions based on regularized tracking data over their non-regularized counterpart.

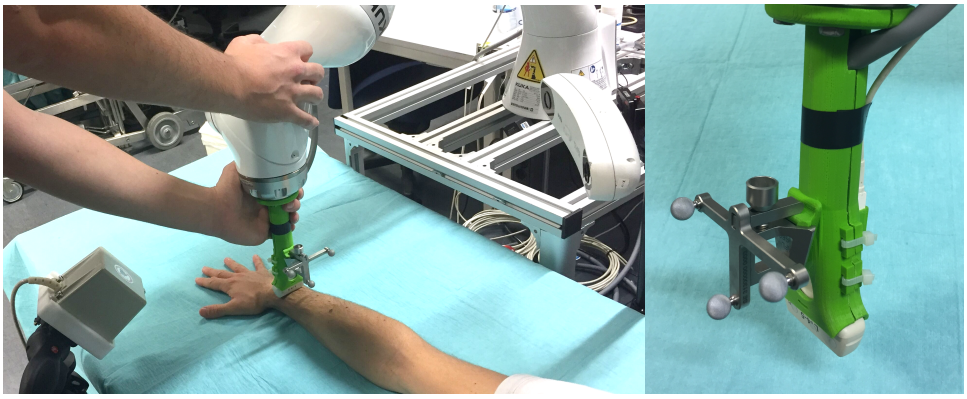


Fig. 6.7. **Experimental setup for the freehand 3D acquisition**. All employed tracking systems are visible in the left picture: the robot (top right), the optical tracking system (bottom right) and the EM field generator (bottom left). The right side depicts a close-up of the custom 3D-printed probe holder, including a support for the optical marker (front side) and an encasing for the external EM sensor.

6.5 Evaluation

In order to offer a comprehensive evaluation of the performance of the regularization algorithm, we performed multiple experimental evaluations. After determining the parameters using the procedure explained in Section 6.4.7, we proceeded to acquire real-world data for further examination.
A setup for ultrasound compounding using optical, electromagnetic and mechanical tracking simultaneously (described in Section 6.5.1 and shown in Fig. 6.7) was assembled, calibrated and used to acquire two distinct datasets. One was a collection of sweeps of a **LEGO phantom**. This particular material has gained in popularity within the community because of the peculiar manufacturing precision [198]. This dataset was acquired by using the robot position controller to perform rectilinear sweeps. A second dataset consists of a freehand sweep of a **human forearm**, acquired by activating the gravity compensation controller of the robotic manipulator.

6.5.1 Setup

The mechanical tracking is implemented with the **robot** used in Chapter 5, i.e. a KUKA iiwa R800. A custom 3D printed mount is designed to hold the US probe, an EM sensor and a marker for the optical tracking system. The **probe** is a L14-5 linear device, connected to an UltraSonix RP machine (Ultrasonix, MA, USA) as in Chapter 5. The probe is equipped with an integrated EM sensor, connected to an NDI Ascension **electromagnetic tracking** system (driveBAY2, Northern Digital Inc., Waterloo, Ontario, Canada). The additional Model 800 sensor is connected to the same EM system, as well as a mid-range field generator. The **optical tracking** system adopted is an NDI Polaris Vicra. The whole setup is depicted in Fig. 6.7.

The 3D printed probe holder is designed to maximize the distance between the EM sensors and the robot. An empirical comparison between the output of the different tracking systems excluded significant interferences.

The ultrasound machine is connected to the robot control workstation via Ethernet, as described in the previous Chapter. The EM tracking system is connected to the same computer over USB. The tracking data is polled from the EM and optical devices by dedicated ROS processes, and forwarded to the tf subsystem. The mechanical tracking data is directly available by using iiwa_stack. The US images were received from the Plus OpenIGTLink server and passed over to a ROS topic. This allowed to record the whole set of images and tracking data using a single rosbag process with on-the-fly compression. As a result, the whole data is available as a single file for simplified storage and later manipulation via Python scripts.

Said data processing is necessary to extract the data from the rosbag file, optionally process it with our regularization method, and then either convert it to the format expected from the compounding software, extract statistics or create visualizations. Ultrasound 3D reconstruction is performed using a backward warping methodology used in the context of previous work within the research laboratory [70]. While the program is capable of multiple and advanced compounding methods, in this thesis mean-compounding is employed because of its widespread utilization in other research contributions.

6.5.2 Calibration

Similar to the previous chapter, the setup employed for all evaluations also requires spatial and temporal calibration methods. Each tracking system needs to be spatially calibrated with respect to the US probe, and in this case we need an accurate temporal calibration because of the great potential impact of minimal lag on the outcome of the compounding.

Based on the integrated EM sensor of the L14-5 probe, the spatial calibration procedure can be simplified significantly. The probe is indeed pre-calibrated by the manufacturer, and the relative position of the integrated EM sensor and the image ${}^{image\_i}_{integrated}T$ can be obtained from the Plus server at runtime, for the current depth settings (as already discussed in Section 5.4.5).

Ideally, the tracking systems should agree on the position of the US image in space, that is:

$$ {}^{image\_i}_{base}T = {}^{image\_e}_{base}T = {}^{image\_r}_{base}T = {}^{image\_o}_{base}T. \tag{6.18} $$

However, this is not true in practice across the whole workspace and while the probe is moving. In order to eliminate the effects of spatially variant tracking accuracies of the whole system, we assume that Eq. 6.18 holds true if a number of transforms are acquired at steady state and averaged subsequently. Thus, postulating that the displacement between each tracking endpoint ($\{integrated\}$ for the integrated EM sensor, $\{external\}$ for the additional sensor attached to the probe mount, $\{flange\}$ for the robot and $\{marker\}$ for the optical system) and the US image does not change due to the rigidity of the probe mount, we can sample

the transformation between the position of the image as reported by the EM tracking system and the position of each other sensor to obtain their spatial calibration matrices (${}^{image\_e}_{external}T$, ${}^{image\_r}_{flange}T$, ${}^{image\_o}_{marker}T$ respectively).

For simplicity, we express all positions with respect to the base of the robot. Performing a **hand-eye calibration** (also as discussed in Section 5.4.5) between the EM system and the robot allows us to find their relative position ${}^{EM}_{base}T$. With this knowledge, we map the output of the EM system into the robot base's frame:

$$ {}^{image\_i}_{base}T = {}^{EM}_{base}T \, {}^{integrated}_{EM}T \, {}^{image\_i}_{integrated}T, \tag{6.19} $$

where ${}^{EM}_{base}T$ is the hand-eye calibration transform, ${}^{integrated}_{EM}T$ is the output of the EM system for the integrated sensor, and ${}^{image\_i}_{integrated}T$ is the pre-calibration matrix as reported by the Plus toolkit.

Consequently, if the transformation between the robot flange and the image is sampled appropriately, the spatial calibration for the robot is obtained by:

$$ {}^{image\_r}_{flange}T = {}^{base}_{flange}T \, {}^{EM}_{base}T \, {}^{integrated}_{EM}T \, {}^{image\_i}_{integrated}T, \tag{6.20} $$

where ${}^{flange}_{base}T$ is the Forward Kinematic of the robot, ${}^{EM}_{base}T$ is the hand-eye calibration with the EM system, ${}^{integrated}_{EM}T$ is the position of the integrated sensor as reported by the EM system and ${}^{image\_i}_{integrated}T$ is the pre-calibration matrix. Once this quantity is known, we can use it to derive the position of the US image according to the mechanical tracking for other poses of the robot

$$ {}^{image\_r}_{base}T = {}^{flange}_{base}T \, {}^{image\_r}_{flange}T, \tag{6.21} $$

where ${}^{flange}_{base}T$ is the Forward Kinematic of the robot and the latter term is defined by the Equation above.

The same operation can be applied to the additional EM sensor, to which the $\{external\}$ reference frame is attached. At a steady state, we sample the transformation

$$ {}^{image\_e}_{external}T = {}^{EM}_{external}T \, {}^{integrated}_{EM}T \, {}^{image\_i}_{integrated}T. \tag{6.22} $$

Again, we can use this calibration to obtain the pose of the US image with respect to the base of the robot according to the external EM sensor

$$ {}^{image\_e}_{base}T = {}^{EM}_{base}T \, {}^{external}_{EM}T \, {}^{image\_i}_{external}T. \tag{6.23} $$

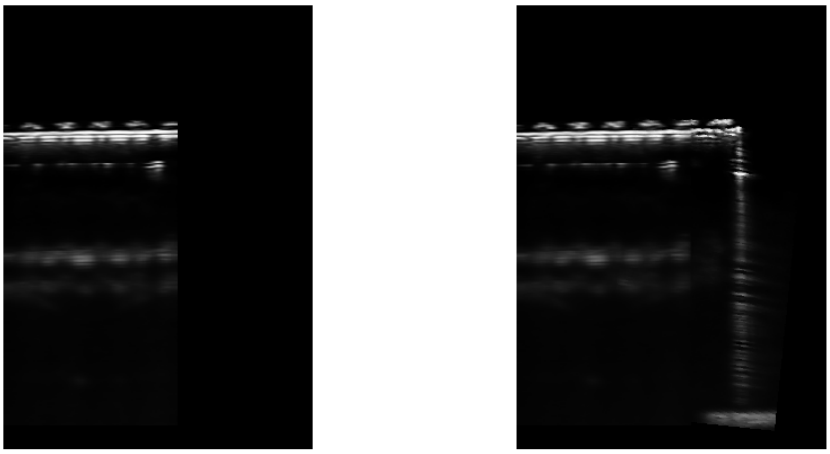


Fig. 6.8. Verification of spatial ultrasound calibration. The LEGO phantom is scanned by moving the robot linearly in one direction, twisting the probe, and traveling back. As a result, the images acquired in the two phases of the acquisition overlapped partially. The slice on the left is from a region where only the first, longer sweep is visible; the right slice contains overlap. It can be noted that there are no artifacts are visible to the naked eye. In case of faulty calibration, a misalignment between the two sweeps are clearly identifiable thanks to this simple test.

After hand-eye calibration between the optical tracking system and the robot, we can sample the relative US spatial calibration matrix

$$ {}^{image\_o}_{marker}T = {}^{optical}_{marker}T \, {}^{base}_{optical}T \, {}^{EM}_{base}T \, {}^{integrated}_{EM}T \, {}^{image\_i}_{integrated}T, \tag{6.24} $$

so that we can obtain the position of the US image according to the optical tracking system:

$$ {}^{image\_o}_{base}T = {}^{optical}_{base}T \, {}^{marker}_{optical}T \, {}^{image\_o}_{marker}T. \tag{6.25} $$

The hand-eye calibration is performed with the same software used in the previous chapter. An additional script is employed to extract the spatial calibration as described here from a simple recording of tracking data of the system in a steady state after hand-eye calibration.

In order to verify the correctness of the calibration, an empirical validation is performed by acquiring a volume of the LEGO phantom with a tilted sweep. If the calibration were faulty, there would be a noticeable disagreement between the two halves of the compounded volume; as it can be seen by visual inspection of Fig. 6.8, this is not the case for the experiments presented in the following sections. Also considering that we would compare volumes reconstructed from the same tracking data (before and after regularization), and that our method does not consider the content of the images, we concluded that the evaluation would not be affected significantly by the quality of the spatial calibration.

After successful hand-eye calibration, the **temporal** mismatch between each tracking system and the US image stream had to be estimated carefully. There are free softwares available for

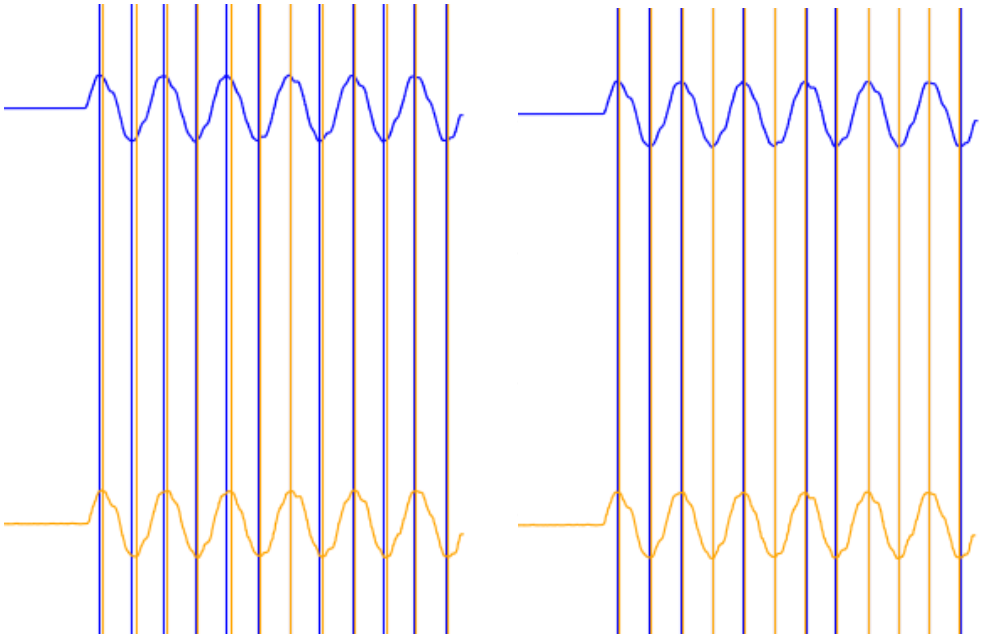


Fig. 6.9. **Temporal calibration of US and tracking devices.** In order to find the temporal offset between the US image and tracking pose streams, the robot is programmed to move in the vertical direction in a sinusoidal fashion while keeping the US probe immersed in a container filled with water. The bottom of the container is segmented from the US image and its vertical coordinate is plotted; the same is done for the vertical component of each tracking stream. The extrema of each plot are found with an automated script, as well as the offset that minimized the average distance between the respective extrema. The offsets are saved and used to synchronize the data throughout the experiments. In this picture, the vertical z coordinate as reported during the temporal calibration procedure by the mechanical tracking system (orange) and derived by the segmentation of the US image (blue) are plotted with respect to the global clock. It is possible to see how the distance between the stationary points is greatly decreased after applying the computed offset (right) with respect to before (left).

such task, but as for the case of spatial calibration they would have required to integrate them with the robotic software used to derive mechanical tracking data. Instead, an analogous approach is quickly developed in-house to provide reliable temporal calibration for the system presented in this chapter.

The robot is programmed to move the probe in the vertical direction in a sinusoidal pattern, while immersed in a water container. The bottom of the container could be directly segmented in the US image via thesholding, since it appeared as a straight line in the image. However, in order to prevent reflection artifacts it is necessary to manually set the US machine to employ a single focal point during B-mode image formation.

As a result, it is possible to correlate the movement along the z axis of the robot base of each tracking system with the vertical movement of the center of mass of the biggest connected region in the US image (corresponding to the container bottom). After processing the streams with Savitzky-Golay interpolation [163] to avoid spurious inflection points, the extrema of the sinusoidal waves could be extracted and aligned. The time displacement between the point series corresponded to the lag between the devices. This process is represented in Fig. 6.9.

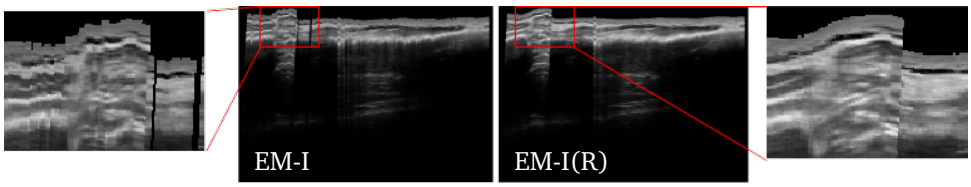


Fig. 6.10. **Effect of tracking regularization on image quality: preservation of discontinuity.** The benefits of our approach over smoothing methods can be noted in particular in correspondence of sudden changes in the direction of the probe trajectory. This can happen when the probe slides over a discontinuity in the tissues, e.g. a vessel. The picture shows how our method manages to subtract the noise while maintaining anatomical detail. It is visible how the tissues underwent greater compression on the right side of the image.

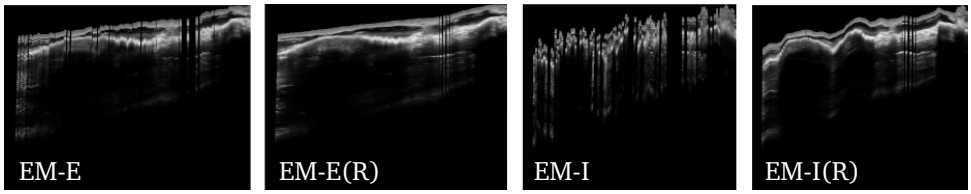


Fig. 6.11. **Effect of tracking regularization on image quality: extreme degradation.** Our regularization method proved to be robust to extreme cases. The volume from which the cross-section depicted in this figure was extracted was acquired in the proximity of a large metallic object. While the method was not designed to compensate for the large scale distortion of the EM tracking introduced by the metalic object, its regularizing effect is noticeable in terms of perceived quality.

6.5.3 Qualitative Evaluation: Compounding of Real Data

After calibrating the system as described in the previous section, the two datasets of the LEGO phantom and a human forearm were acquired and reconstructed. Considering the qualitative evaluation as indicated in Figures 6.12 and 6.13, a clear improvement can be observed by visual inspection of the cross-sections of the volumes compounded using different tracking systems with and without regularization. The value of the proposed approach becomes in particular clear for the integrated EM tracking. It is worth noting how the method does not introduce an artificial smoothing of the trajectories and hence maintained detail even under sharp changes in the direction of the probe (Figures 6.12, 6.13, represent the general cases, while pathologic situations are studied in Figures 6.10, 6.11).

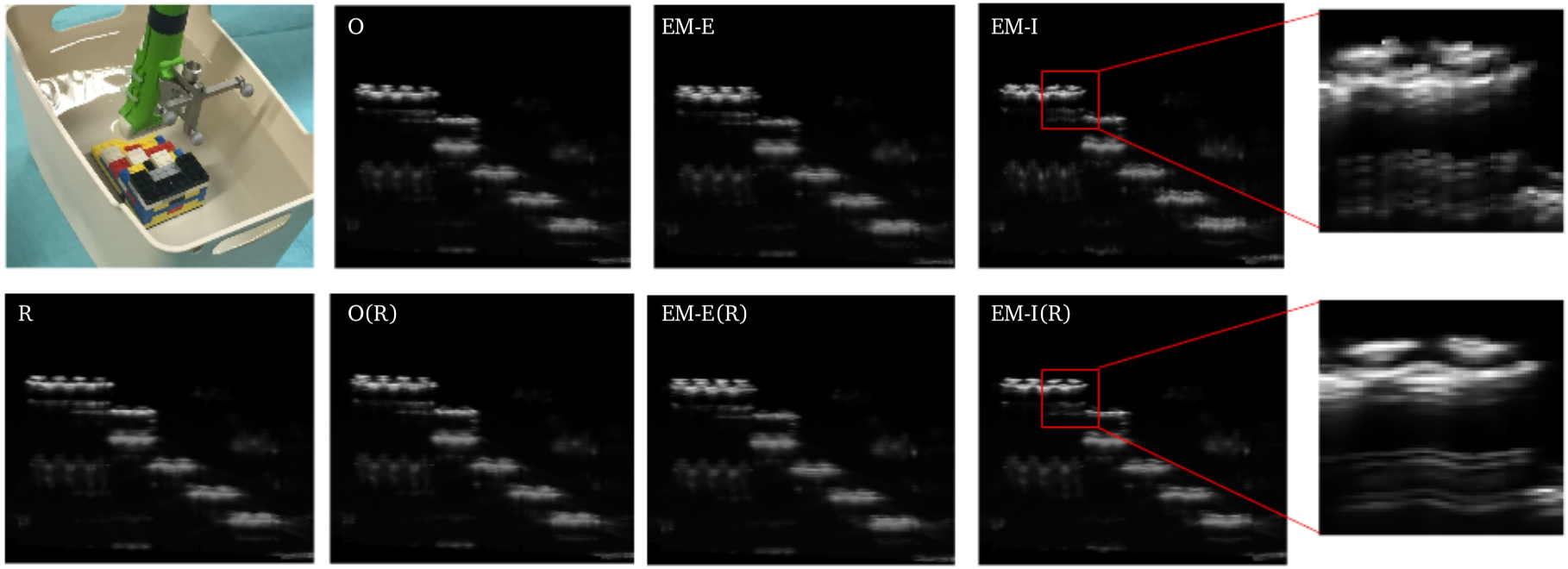


Fig. 6.12. **Effect of tracking regularization on image quality: evaluation on a LEGO phantom.** The picture in the top left corner shows the phantom in a water bath, and the US probe with the markers held by the robot. Directly below it, a section of the volume compounded according to mechanical tracking (R) is visualized. The three successive columns contain the compoundings obtained through optical tracking (O), the external EM sensor (EM-E) and the integrated EM sensor (EM-I) respectively; the volumes in the bottom row of these three columns were reconstructed after tracking data regularization. The improvement is clearly noticeable in the magnifications, which refer to the integrated EM sensor. The ondulation that can be seen in the EM volumes may be due to the action of the robot motors, which were running in this case; distortion compensation is not in the scope of this work, but it is worth noting that our method did not perform an overregularization.

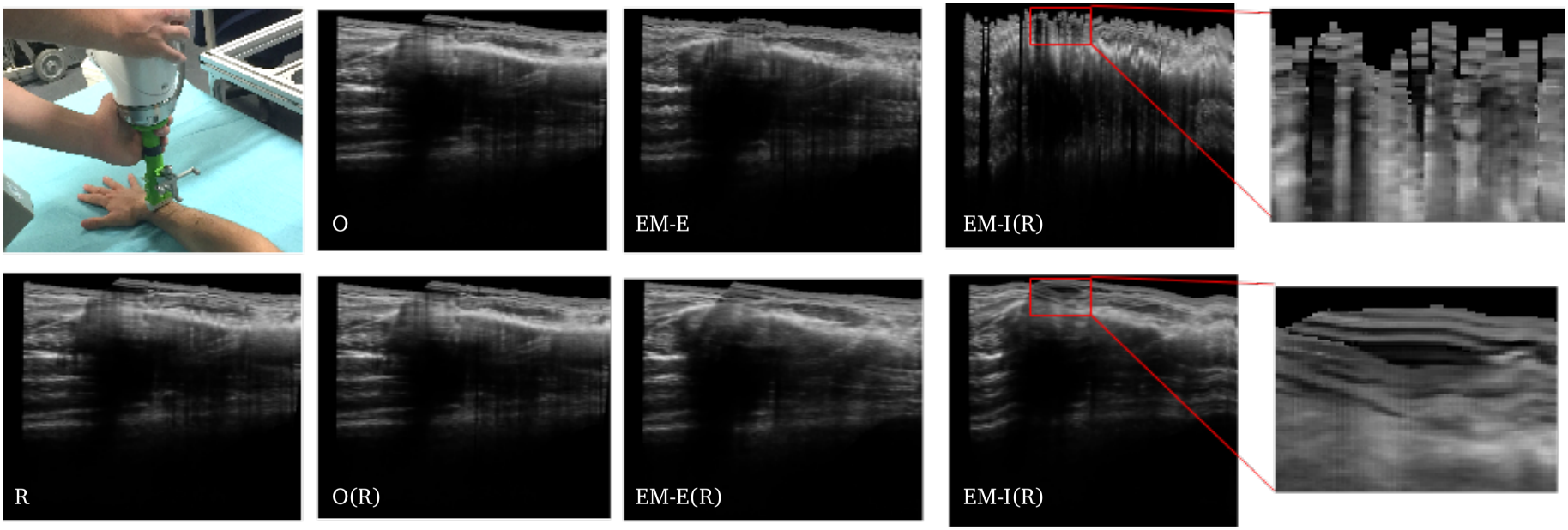


Fig. 6.13. **Effect of tracking regularization on image quality: evaluation on human anatomy.** As in the previous Figure, the setup can be seen in the top left corner. The picture in the bottom left was obtained by compounding the US frames on the basis of mechanical tracking data (R). The disposition of the other three columns is also the same as in the previous case: optical tracking (O), external EM sensor (EM-E), integrated EM sensor (EM-I) from left to right; non-regularized data on the top, regularized in the bottom row. The magnification makes the improvement clear once again, despite the heavy noise present in the tracking data provided by the integrated EM sensor.

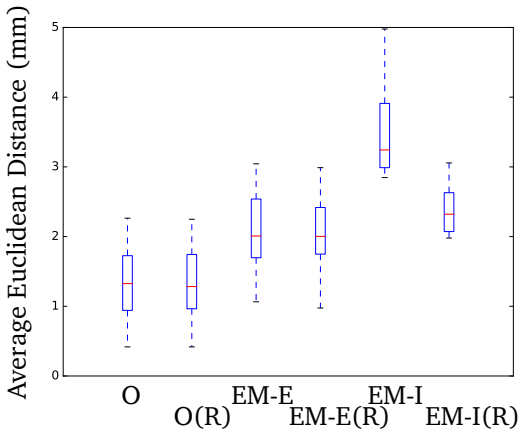


Fig. 6.14. **Effect of regularization on real-world tracking data.** This plot shows the average Euclidean distance between each point of the trajectory recorded by the mechanical tracking system and the respective point in the other tracking system data (i.e. the closest in time of acquisition). The optical tracking system (**O**) does not benefit significantly from the regularization (**O(R)**), but it also not harmed by it. The external EM sensor (**EM-E**) data is slightly improved after regularization (**EM-E**). However, the integrated EM sensor (**EM-I**) benefits significantly from the application of our method (**EM-I(R)**), getting remarkably close to the performance of the external sensor. It can be theorized that the discrepancy between the EM and the mechanical tracking may be due to a distortion effect. However, it is also worth noting that the disagreement between EM and mechanical tracking is higher but of the same order of magnitude of that between mechanical and optical tracking.

6.5.4 Quantitative Evaluation: Measurement of Impact on Real Tracking Data

A quantitative evaluation shows how our regularization method brought the integrated EM tracking to a closer agreement to our reference systems, i.e. optical and mechanical tracking. The trajectories relative to the compoundings shown in the previous section were used to extract statistics reflecting the tracking quality. Since we are interested in the local characteristics of the recorded trajectories, we registered each trajectory via the Iterative Closest Point (ICP) algorithm [12] to the one acquired via mechanical tracking, and then computed the respective average Euclidean distance.

As it is shown in Fig. 6.14, the application of our denoising method reduced the distance between the EM trajectories and the ones acquired by the robot. Also, the visual comparison of compounded results shows in overall more continuous structures. On the other hand, optical tracking was not subject to a significant improvement. However, it is still worth noting that it was not worsened either, which indicates the absence of over-regularization. Furthermore, the disagreement between the optical and mechanical tracking reminds us that perfect agreement between two tracking systems is never to be expected, since each of them have their own idiosyncrasies and do not perfectly match with the physical "reality". Hence, the average distance between the optical and mechanical tracking acts as a baseline for the maximum performance achievable with contemporary technology.

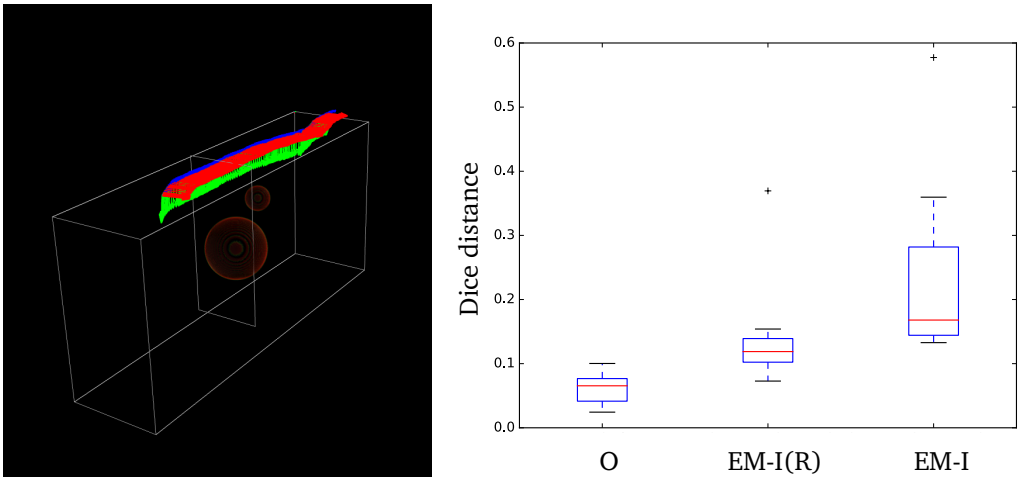


Fig. 6.15. **Quantitative Evaluation of Compounding Quality.** Left: A binary volume, containing zero values with the exception of two spheres, is sampled according to the robot tracking data in order to create a set of synthetic binary US images. The obtained compounding is then compared to the compounding obtained by using other tracking systems, c.f. also Fig. 6.16. Right: Box plot of DICE values computed between the compounding obtained with robotic tracking and the compoundings obtained with optical (O), internal electromagnetic (EM-I) and regularized internal electromagnetic tracking (EM-I(R)), respectively. The improvement of the quality achievable with EM tracking is evident when comparing the performance before denoising (right-most column) to the one obtained after regularization (middle column). While the performance of optical tracking is best (left-most column), the proposed method improves the resulting average quality, but most notably its consistency, in comparison to the outcome obtained without regularization.

6.5.5 Quantitative Evaluation: Measurement of Impact on Ultrasound Compounding

As we discussed in Section 6.2, most of the available literature about error correction for ultrasound compounding addresses distortion correction for electromagnetic tracking. Since distortion leads to deformations of the image on a global scale, in most of the cases the method performance is evaluated by scanning fiducial phantoms of known geometry and successively compare the measurement of salient features in the compounded volume to the actual dimensions (e.g. the distance between two points).

Such approaches do not reflect the performance of our method, since we address local artifacts in the image which do not involve large-scale deformations. Quantitative approaches to measure *image quality* are indeed difficult to find in literature, in large part due to the low signal-to-noise ratio of ultrasound B-mode scans. These images present an elevated amount of speckle artifacts, which are similar in appearance to *salt-and-pepper noise* and make standard approaches for image comparison like Sum of Absolute Differences (SAD) inconclusive.

In virtue of this, we rely on an original method to quantify the impact of tracking noise on ultrasound compounding.
The reasoning is that since our method only affects the positioning of the US slices in space, their content is actually irrelevant and we can replace it with something that would lead to meaningful image statistics after compounding. In order to generate such images, a synthetic

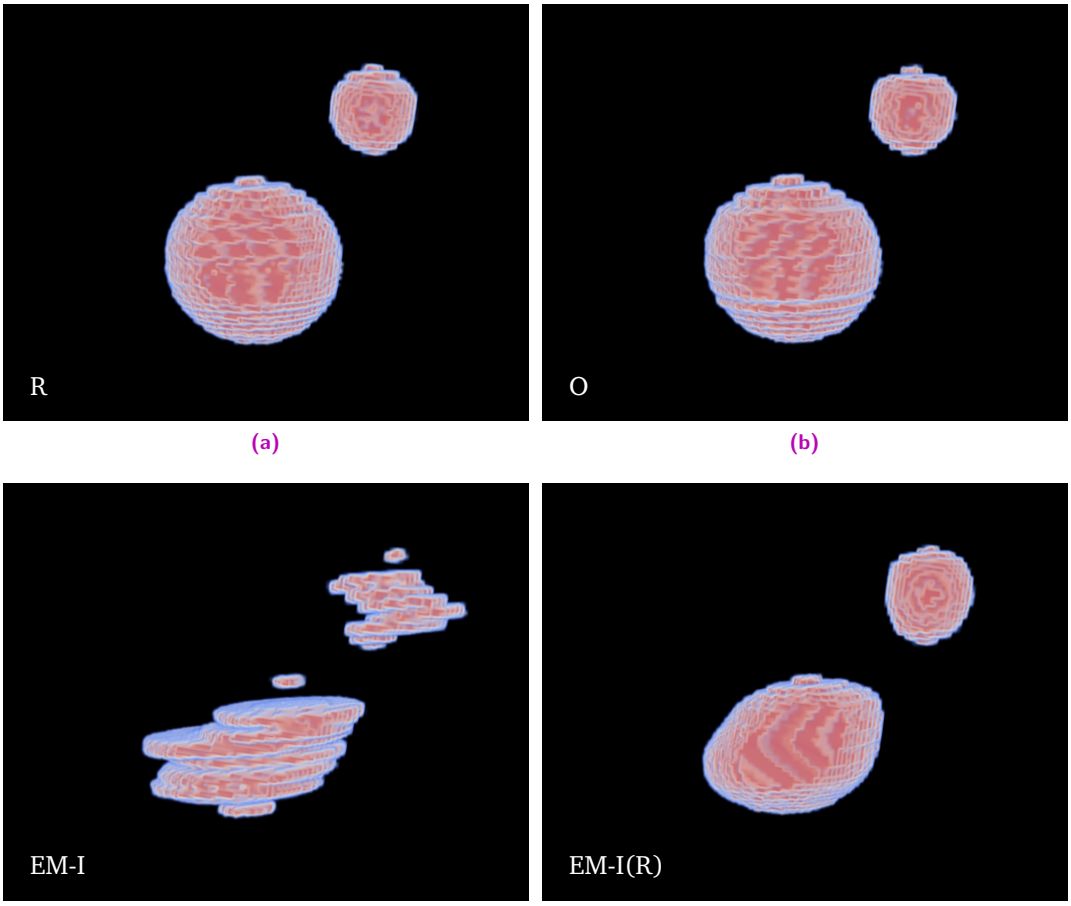


Fig. 6.16. **Results of ultrasound compounding reconstruction of synthetic data.** The data is generated by reslicing the synthetic volume according to mechanical tracking data. Unsurprisingly, the volume reconstruction obtained according to the same tracking information lead to the best perceived quality (a). Optical tracking agreed almost perfectly with mechanical tracking, and the result of the reconstruction confirms this observation (b). The data from the integrated EM sensor turned into the worst results (c). However, our regularization method is still able to lead to an intelligible result even in the worst case among all our acquisitions, which is the one reported in this figure.

3D binary labelmap was generated. The bounding box of the tracking data was filled with a black background, from which a couple of white spheres were subtracted. For each pose of the US probe in the tracking stream, a 2D slice was computed by reslicing the volume accordingly (the process is depicted in Fig. 6.15). The result was a "synthetic ultrasound scan" of white spheres on a black background. Given a sufficient sampling density and coverage, using the same tracking data would lead to the original volume. This could be verified by computing the SAD between the compounding and the original data; the expected value would be zero. However, if the position of the slices in space would change, their alignment would not be exact anymore, artifacts could be seen in the image and the SAD (or comparable measurement) with the original volume would be greater than zero. Thus, reconstructing the slices generated with mechanical tracking data according to the poses provided by another tracking system would give us an image on which meaningful image quality statistics can be computed (the volumes relative to the most noisy acquisition are visible in Fig. 6.16).

We translated this thought experiment into practice, and we were able to verify our hypothesis. Application of our regularization method resulted in a significant drop of the Dice distance of the obtained volumetric images to the reference volume.

6.6 Outlook and Future Work

In this chapter we discussed an approach to improve the perceived quality of freehand 3D ultrasound images. With this aim, we presented a regularization method capable of reducing the level of jitter affecting the pose stream produced by a tracking system, without external information from additional sources. In contrast to comparable methods, the adoption an explicit data fidelity term leads the optimization to maximize the smoothness of the result while at the same time minimizing the distance from the input trajectory. This translates to images that are more pleasant and easier to interpret for the operator. However, such effect cannot be measured with methods that are commonly used to evaluate distortion correction approaches. Instead, we use real-world mechanical tracking data to reslice a synthetic binary volume to give origin to a set of synthetic US images which are free of speckle artifacts. After compounding this artificial sweep according to EM tracking data before and after regularization we obtain speckle-free volumes which can be meaningfully compared via voxel-wise similarity measures.

The proposed method addresses the particular issue of perceived quality of ultrasound compounding, which is underrepresented in literature and for which there are no established evaluation strategies. The successful preliminary study run in this context would make a comprehensive comparison to alternative methods such as Moving Average and Extended Kalman Filter an interesting avenue to be explored in future.
Also, as previously mentioned, the integration with a Distortion Correction method would be worth evaluating.
Finally, it would be interesting to investigate if the interpolation between tracking poses for better agreement with ultrasound timestamps would be improved by prior regularization. Such interpolation was not performed here (the closest tracking pose to the US frame was chosen), but such an approach can potentially lead to further improvement in image quality.

Part IV

Conclusion

7 Conclusion and Outlook

In this dissertation we have presented an overview of the advancing field of medical imaging and robotics, and following the arguments of their respective advantages and shortcomings we proposed novel approaches that augment intraoperative imaging. As we discussed in Section 5.7, the introduction of a collaborative robotic assistant enables real-time multimodal imaging within the familiar workflow of ultrasound-guided needle biopsy. The addition of functional information to the anatomical image allows to unequivocally identify the sentinel lymph node among all the visible axillary nodes. The robotic arm works together with the surgeon, moving the gamma camera in space, thus reducing the cognitive burden that is posed by the conventional workflow involving the two imaging devices separately and two subsequent phases of SLN identification and US-guided biopsy. As we mentioned in Section 3.5, this factor can be decisive for the success of the procedure. The real-time nature of the system allows to overcome difficulties posed by the deformation of the soft tissue in the breast and always guarantees a correct image. These factors result in a workflow which can be picked up by beginners in a few seconds and leads to high sensitivity. The conversion of our research system employing a standard industrial manipulator into a prototype making use of a custom designed robot and additional safety features incrementing the awareness of the environment is still a necessary prerequisite for a clinical study capable to prove the viability of SLN biopsy under multimodal guidance for breast cancer staging.
In Section 6.6 we highlighted how the adoption of a Riemannian regularization framework allows to remove jitter from the output of a tracking system, while optimizing for both fidelity to the original input and smoothness of the output at the same time. We discussed how freehand 3D ultrasound can in particular benefit from this effect, given the large adoption of electromagnetic tracking in such systems due to ergonomic concerns and the relationship between tracking jitter and final image quality on a local scale. We performed synthetic and real-world experiments to show the qualitative impact of this technique, and we employed an *ad-hoc* strategy to quantify this effect. Our results show a definite improvement for the challenging scenarios in which the quality of EM tracking degrades sensibly in terms of precision. Conjugated with a complementary distortion correction method, our solution may significantly improve the quality of freehand 3D ultrasound both on a global and a local scale.

Our work fits within the panorama of solutions that technology has made available to support physicians in their work; or, to be more precise, *medical teams*. As we mentioned throughout this text, further progress in the area of Computer Assisted Interventions may profit from a **holistic approach** to research and product development alike, transcending the boundaries between traditional areas of competence. Engineers will need to get in closer contact with surgeons, in order to produce systems that will result in measurable improvement in patient care rather than gadgets. On the other hand, if doctors will find the time and resources to try innovative approaches they may open the door to radical new approaches, capable of surpassing anachronistic limitations by now taken for granted. Traditional "siloed" research

will be optimally integrated into clinical practice by combining resources into interdisciplinary projects, and conventional hospital management may one day be revolutionized by a process-oriented optimization. Hopefully, the result will be a transition from an OR resembling an incoherent patchwork of idiosyncratic technologies to a cohesive, distraction-free assembly of well-integrated devices where the surgeon can focus on healing the patient rather than coping with the equipment.

A recurrent topic regarding the future of medicine is the discussion about the **level of automation** that should be reached. On the diagnostic side, it is debated if machines could completely replace radiologists to achieve superior diagnostic outcome [130], and even surgeons are also periodically proposed as candidates for eventual replacement [25, 219]; such enthusiastic visions are regularly met by a choir of skepticism [218, 214]. While waiting for history to unroll itself, it can be interesting to examine the factors that will exercise their weight in the process and speculate about future developments.

In Section 4.1 we reported the current consensus about the strengths and weaknesses of humans and robots, with the latter showing advantages in terms of **motion** precision and repeatability while also revealing limited flexibility and capability to adapt to unknown situations. We also discussed the successes in automating subtasks in human-in-the-loop setups in Section 4.2.3. On this basis, it could be speculated that an increasing number of routine subtasks could be delegated to autonomous software, capable of optimizing the robot motion according to the geometric configuration as well as to information provided by pre- and intraoperative imaging. As a result, the role of the surgeon could gradually shift to providing the surgical robot with information about *what* to do rather than *how*, and leave the concrete implementation of the plan to the machine. Such approach could be defined as a *declarative* control method, opposed to the current *imperative* schemes where the robot's motion mirrors more or less transparently the input data.

A similar symbiotic extension of human capabilities could be imagined for **memory**. Currently, the surgeon must rely on their own knowledge and experience, acquired from textbooks and attended or performed surgeries. However, there is a limit to the amount of information that a human can memorize, and the speed at which one can learn. It is hence fathomable that the fruits of Surgical Data Science, which we mentioned in Section 3.3.4 could function as a tool to overcome this other human limitation by providing, on demand or proactively, the most opportune information from a virtually endless knowledge base accumulated over time in surgeries all over the world. In this way, the integration of technology into the OR could serve as the foundation of a collective intelligence built by the insight and experience of a pool of surgeons connected over space and time.

The projection of such ideas into the future could lead to a point where artificial intelligence has accumulated so much experience by observing clinical cases and their treatment, that the human contribution in the planning of the intervention could be no more necessary and the future *robot surgeons* could operate completely autonomously. While it is impossible to foresee if this eventuality may be at all reachable or not, it is easy to imagine how long the path to it would be: medicine is typically conservative and resistant to change (may it be the introduction of the endoscope and multimodal imaging as we mentioned in our introduction, or even a simple thermometer [116]). And rightly so, since a physician carries the huge responsibility of the patient's health and, ultimately, life. However, it may be interesting to entertain a parallel with another field where automation has been discussed for more than 40 years: automobile production.

In the beginning of the 1980s Japanese competition in the US market was eroding the profit of local car producers. As a response, General Motors spearheaded the automation efforts of the whole sector, traditionally affine to robotics, with a massive investment plan [49]. The aim was to achieve higher productivity and lower costs, but this was never the case; in fact, these indicators worsened in parallel to the progressive automation of the assembly line. During the same time, Toyota was refining the lean strategy popularized by Womack's bestseller "The Machine that changed the world" [204]. By addressing inefficiencies and wastes in the production process using a holistic approach, by 1986 Toyota was able to produce four times more cars per employee than GM (which had already spent five years in their automation efforts) [49]. This surprising lesson seems to be periodically forgotten but it stays valid as of today, as Tesla's recent attempt at and hasty dismissal of fully automated assembly showed [216, 217].

These events pertain a completely different field than medicine, but their implications are even stronger because of that. A car manufacturing plant is a very controlled environment, where the geometry of the objects of interest is completely known in advance and tasks are repetitive and well defined. Clear success stories in such scenarios seem like a natural prerequisite before approaching tasks so complex, uncertain and critical as surgeries. This may be bound to happen; as in the case of Autonomous Driving, it can be argued that the objective should not be to reach absolutely perfect outcomes (which may be unfeasible), but to perform better than the human alternative. In the meantime, as also suggested by other authors [128, 158], methods involving the integration of the complementary characteristics of humans and machines may provide an incremental way of reaching such objective while achieving increasing performance.
However, it is important to remember that increasingly complex technology is not the only mean available to us for improving clinical quality and outcome. An amazing counterexample is provided by the "Paperfuge" [13], a 0.20$ worth device consisting in an elastic and a paper disk, capable of replacing the very expensive blood centrifuges needed to diagnose blood diseases such as malaria. This innovation, as low-tech as disruptive, may revolutionize healthcare in developing countries where such illnesses are most common. Another low-cost solution is represented by the Surgical Safety Checklist [68], a protocol developed to minimize human error during surgical procedures. Its adoption has been reported to have reduced both mortality and complication rates by half, without introducing new devices but just adopting the process standardization already common in other mission-critical industries.
All in all, we can still be confident that the human factor will continue being the principal component of progress, and that the strength, accuracy and resilience of machines will be precious *tools* supporting human creativity in breaking new barriers on the path to a better future.

Part V

Appendix

List of Authored and Co-authored Publications

A

2019

[42] **Marco Esposito**, Christoph Hennersperger, Rüdiger Göbl, Laurent Demaret, Martin Storath, Nassir Navab, Maximilian Baust, and Andreas Weinmann. "Total Variation Regularization of Pose Signals with an Application to 3D Freehand Ultrasound". *IEEE Transactions on Medical Imaging (TMI)*, 2019.

2018

[121] Fausto Milletari, Nicola Rieke, Maximilian Baust, **Marco Esposito**, and Nassir Navab. "CFCM: Segmentation via Coarse to Fine Context Memory". *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2018, Granada, Spain*.

2017

[212] Oliver Zettinig, Benjamin Frisch, Salvatore Virga, **Marco Esposito**, Anna Rienmüller, Bernhard Meyer, Christoph Hennersperger, Yu-Mi Ryang and Nassir Navab. "3D Ultrasound Registration-based Visual Servoing for Neurosurgical Navigation". *International Journal of Computer Assisted Radiology and Surgery, 2017*.

2016

[22] Benjamin Busam, **Marco Esposito**, Benjamin Frisch and Nassir Navab. "Quaternionic Upsampling: Hyperspherical Techniques for 6 DoF Pose Tracking". *IEEE International Conference on 3DVision, 2016, Stanford, USA*.

[195] Salvatore Virga, Oliver Zettinig, **Marco Esposito**, Karin Pfister, Benjamin Frisch, Thomas Neff, Nassir Navab and Christoph Hennersperger. "Automatic force-compliant robotic ultrasound screening of abdominal aortic aneurysms". *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2016, Daejeon, South Korea*.

[41] **Marco Esposito**, Benjamin Busam, Christoph Hennersperger, Julia Rackerseder, Nassir Navab, and Benjamin Frisch. "Multimodal US–gamma imaging using collaborative robotics for cancer staging biopsies". *International Journal of Computer Assisted Radiology and Surgery, 2016*.

[213] Oliver Zettinig, Bernhard Fuerst, Risto Kojcev, **Marco Esposito**, Mehrdad Salehi, Wolfgang Wein, Julia Rackerseder, Edoardo Sinibaldi, Benjamin Frisch and Nassir Navab. "Toward real-time 3D ultrasound registration-based visual servoing for interventional navigation". *IEEE International Conference on Robotics and Automation (ICRA), 2016, Stockholm, Sweden*.

2015

[103] Roberto Londei, **Marco Esposito**, Benoit Diotte, Simon Weidert, Ekkehard Euler, Peter-Helmut Thaller, Nassir Navab and Pascal Fallavollita. "Intraoperative augmented reality in distal locking". *International Journal of Computer Assisted Radiology and Surgery, 2015*.

[21] Benjamin Busam, **Marco Esposito**, Simon Che'Rose, Nassir Navab and Benjamin Frisch. "A Stereo Vision Approach for Cooperative Robotic Movement Therapy". *IEEE International Conference on Computer Vision Workshop (ICCVW), 2015, Santiago, Chile*.

[40] **Marco Esposito**, Benjamin Busam, Christoph Hennersperger, Julia Rackerseder, An Lu, Nassir Navab, and Benjamin Frisch. "Cooperative robotic gamma imaging: Enhancing us-guided needle biopsy". *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015, Munich, Germany*.

2014

[104] Roberto Londei, **Marco Esposito**, Benoit Diotte, Simon Weidert, Ekkehard Euler, Peter-Helmut Thaller, Nassir Navab and Pascal Fallavollita. "The Augmented Circles: A Video-Guided Solution for the Down-the-Beam Positioning of IM Nail Holes". *Information Processing in Computer-Assisted Interventions (IPCAI), 2015, Barcelona, Spain*.

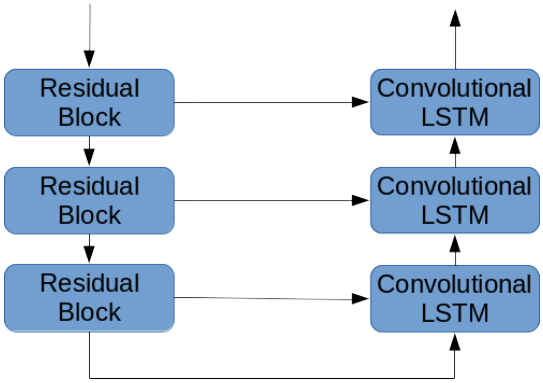
[53] José Gardiazabal, **Marco Esposito**, Philipp Matthies, Aslı Okur, Jakob Vogel, Silvan Kraft, Benjamin Frisch, Tobias Lasser and Nassir Navab. "Towards personalized interventional SPECT-CT imaging". *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2014, Boston, USA*.

B

Abstracts of Publications not Discussed in this Thesis

CFCM: Segmentation via Coarse to Fine Context Memory

Fausto Milletari, Nicola Rieke, Maximilian Baust, Marco Esposito, Nassir Navab



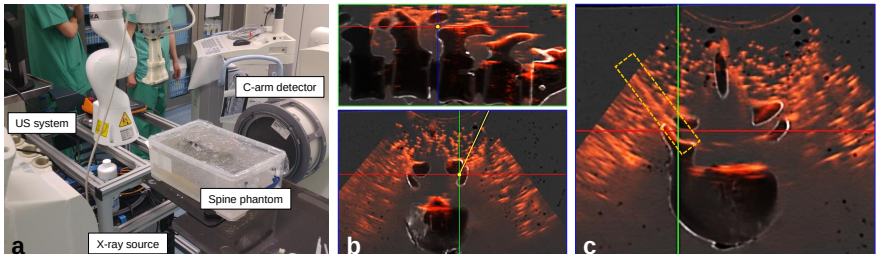
Network Architecture: ResNet residual blocks are organized in an Encoder/Decoder fashion, with skip connections between them. Each residual block in the encoding branch extracts features at a coarser scale, hence at a more global abstraction level. The result is fed to the decoder, which is composed of LSTM blocks.

Recent neural-network-based architectures for image segmentation make extensive usage of feature forwarding mechanisms to integrate information from multiple scales. Although yielding good results, even deeper architectures and alternative methods for feature fusion at different resolutions have been scarcely investigated for medical applications. In this work we propose to implement segmentation via an encoder-decoder architecture which differs from any other previously published method since (i) it employs a very deep architecture based on residual learning and (ii) combines features via a convolutional Long Short Term Memory (LSTM), instead of concatenation or summation. The intuition is that the memory mechanism implemented by LSTMs can better integrate features from different scales through a coarse-to-fine strategy; hence the name Coarse-to-Fine Context Memory (CFCM). We demonstrate the remarkable advantages of this approach on two datasets: the Montgomery county lung segmentation dataset, and the EndoVis 2015 challenge dataset for surgical instrument segmentation.

Medical Image Computing and Computer Assisted Intervention – MICCAI 2018.
Reprinted with permission from Springer.
DOI: https://doi.org/10.1007/978-3-030-00937-3_76

3D Ultrasound Registration-based Visual Servoing for Neurosurgical Navigation

Oliver Zettinig, Benjamin Frisch, Salvatore Virga, Marco Esposito, Anna Rienmüller, Bernhard Meyer, Christoph Hennersperger, Yu-Mi Ryang, Nassir Navab



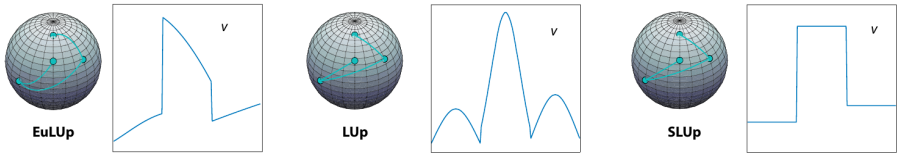
(a) System setup: the robot holds a wobbler 3D Ultrasound probe via a custom 3D printed mount, including a pre-calibrated needle guide. (b) Intervention planning: the physician selects the desired trajectory for the needle within the pre-operative CT data, registered to an intra-operative compounded volume. (c) Accuracy verification: the agreement between the plan and the effective needle trajectory can be assessed in the Ultrasound image and with an additional X-Ray image (acquired only for verification purposes).

Purpose We present a fully image-based visual servoing framework for neurosurgical navigation and needle guidance. The proposed servo-control scheme allows for compensation of target anatomy movements, maintaining high navigational accuracy over time, and automatic needle guide alignment for accurate manual insertions.
Method Our system comprises a motorized 3D ultrasound (US) transducer mounted on a robotic arm and equipped with a needle guide. It continuously registers US sweeps in real time with a pre-interventional plan based on CT or MR images and annotations. While a visual control law maintains anatomy visibility and alignment of the needle guide, a force controller is employed for acoustic coupling and tissue pressure. We validate the servoing capabilities of our method on a geometric gel phantom and real human anatomy, and the needle targeting accuracy using CT images on a lumbar spine gel phantom under neurosurgery conditions.
Results Despite the varying resolution of the acquired 3D sweeps, we achieved direction-independent positioning errors of 0.35 ± 0.19 mm and 0.61°± 0.45°, respectively. Our method is capable of compensating movements of around 25 mm/s and works reliably on human anatomy with errors of 1.45 ± 0.78 mm. In all four manual insertions by an expert surgeon, a needle could be successfully inserted into the facet joint, with an estimated targeting accuracy of 1.33 ± 0.33 mm, superior to the gold standard. Conclusion: The experiments demonstrated the feasibility of robotic ultrasound-based navigation and needle guidance for neurosurgical applications such as lumbar spine injections.

International Journal of Computer Assisted Radiology and Surgery. September 2017, Volume 12, Issue 9, pp 1607–1619.
Reprinted with permission from Springer.
DOI: https://doi.org/10.1007/s11548-017-1536-2

Quaternionic Upsampling: Hyperspherical Techniques for 6 DoF Pose Tracking

Benjamin Busam, Marco Esposito, Benjamin Frisch, Nassir Navab



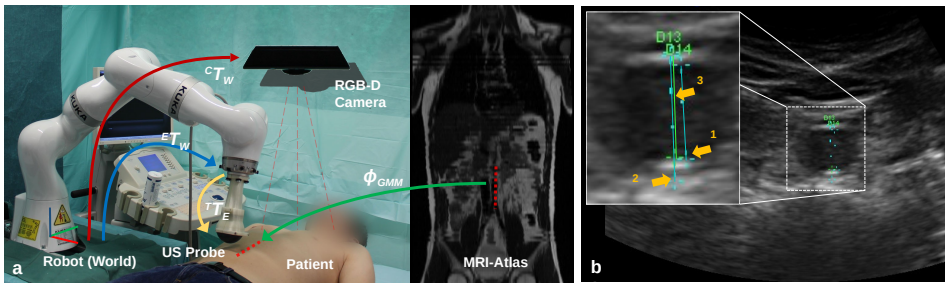
Effect of the interpolation method on the resulting angular velocity. © 2016 IEEE

Fast real-time tracking is an integral component of modern 3D computer vision pipelines. Despite their advantages in accuracy and reliability, optical trackers suffer from limited acquisition rates depending either on intrinsic sensor capabilities or physical limitations such as exposure time. Moreover, data transmission and image processing produce latency in the pose stream. We introduce quaternionic upsampling to overcome these problems. The technique models the pose parameters as points on multidimensional hyperspheres in (dual) quaternion space. In order to upsample the pose stream, we present several methods to sample points on geodesics and piecewise continuous curves on these manifolds and compare them regarding accuracy and computation efficiency. With the unified approach of quaternionic upsampling, both interpolation and extrapolation in pose space can be done by continuous linear variation of only one sampling parameter. Since the method can be implemented rather efficiently, pose rates of over 4 kHz and future pose predictions with an accuracy of 128 μm and 0.5°are possible in real-time. The method does not depend on a special tracking algorithm and can thus be used for any arbitrary 3 DoF or 6 DoF rotation or pose tracking system.

Proceedings of the Fourth International Conference on 3D Vision (3DV) 2016. Reprinted with permission. © 2016 IEEE
DOI: https://doi.org/10.1109/3DV.2016.71

Automatic force-compliant robotic ultrasound screening of abdominal aortic aneurysms

Salvatore Virga, Oliver Zettinig, Marco Esposito, Karin Pfister, Benjamin Frisch, Thomas Neff, Nassir Navab, Christoph Hennersperger



(a) System Overview: the robot holding the ultrasound probe is co-registered with an RGB-D camera, which allows to capture the outline of the patient's skin. The resulting point cloud is deformably registered with the profile of an MRI Atlas, so that a pre-operative plan can be specified on the Atlas and carried out on the real patient. (b) Clinical Measurement: the robot performs a freehand ultrasound sweep autonomously, and the resulting 3D image can be manually or automatically annotated to measure the diameter of the aorta.

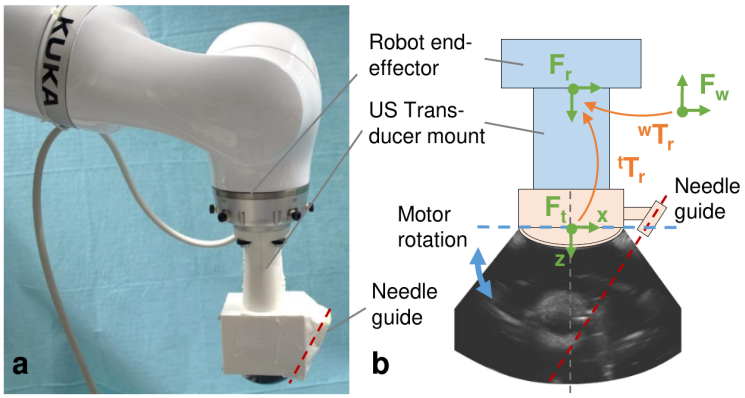
Ultrasound (US) imaging is commonly employed for the diagnosis and staging of abdominal aortic aneurysms (AAA), mainly due to its non-invasiveness and high availability. High inter-operator variability and a lack of repeatability of current US image acquisition impair the implementation of extensive screening programs for affected patient populations. However, this opens the way to a possible automation of the procedure, and recent works have exploited the use of robotic platforms for US applications, both in diagnostic and interventional scenarios. In this work, we propose a system for autonomous robotic US acquisitions aimed at the quantitative assessment of patients' vessel diameter for abdominal aortic aneurysm screening. Using a probabilistic measure of the US quality, we introduce an automatic estimation of the optimal pressure to be applied during the acquisition, and an online optimization of the out-of-plane rotation of the US probe to maximize the visibility of the aorta. We evaluate our method on healthy volunteers and compare the results to manual acquisitions performed by a clinical expert, demonstrating the feasibility of the presented system for AAA screening.

Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).

DOI: https://doi.org/10.1109/IROS.2016.7759101

Toward real-time 3D ultrasound registration-based visual servoing for interventional navigation

Oliver Zettinig, Bernhard Fuerst, Risto Kojcev, Marco Esposito, Mehrdad Salehi, Wolfgang Wein, Julia Rackerseder, Edoardo Sinibaldi, Benjamin Frisch, Nassir Navab



(a) Picture of the robot holding the US probe. (b) Schematic overview of the reference frames involved in the system calibration. © 2016 IEEE

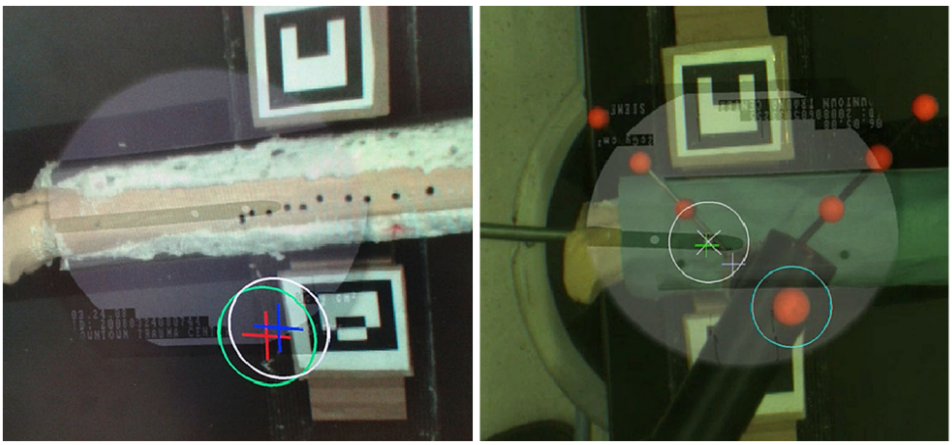
While intraoperative imaging is commonly used to guide surgical interventions, automatic robotic support for image-guided navigation has not yet been established in clinical routine. In this paper, we propose a novel visual servoing framework that combines, for the first time, full image-based 3D ultrasound registration with a real-time servo-control scheme. Paired with multi-modal fusion to a pre-interventional plan such as an annotated needle insertion path, it thus allows tracking a target anatomy, continuously updating the plan as the target moves, and keeping a needle guide aligned for accurate manual insertion. The presented system includes a motorized 3D ultrasound transducer mounted on a force- controlled robot and a GPU-based image processing toolkit. The tracking accuracy of our framework is validated on a geometric agar/gelatin phantom using a second robot, achieving positioning errors of on average 0.42 ± 0.44 mm. With compounding and registration runtimes of up to total around 550 ms, real-time performance comes into reach. We also present initial results on a spine phantom, demonstrating the feasibility of our system for lumbar spine injections.

Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA).

DOI: https://doi.org/10.1109/ICRA.2016.7487226

Intra-operative augmented reality in distal locking

Roberto Londei, Marco Esposito, Benoit Diotte, Simon Weidert, Ekkehard Euler,
Peter-Helmut Thaller, Nassir Navab, Pascal Fallavollita



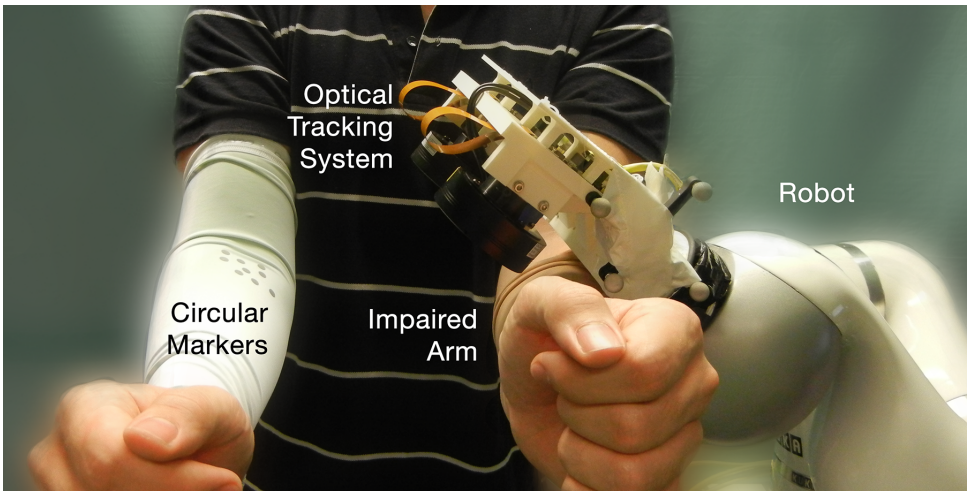
On the left, the Augmented Reality guidance system for Down-the-Beam positioning. The 6 Degrees of Freedom pose of the intramedullary nail is derived by a single X-ray image and used for intra-operative guidance in repositioning the X-ray machine to a Down-the-Beam position. On the right hand side, another AR guidance system for drilling the hole for the IM nail screw. The surgeon can place the drill and perform the boring without further X-ray acquisitions. The whole workflow can be carried out with a single X-ray shot, as opposed to the multiple acquisitions taking place with conventional workflows.

Purpose To design an augmented reality solution that assists surgeons during the distal locking of intramedullary nailing procedures.
Method Traditionally, the procedure is performed under X-ray guidance and requires a significant amount of time and radiation exposure. To absolve these complications, we propose video guidance that allows surgeons to achieve both the down-the-beam position of the intramedullary nail and its subsequent locking. For the down-the-beam position, the IM nail pose in X-ray is calculated using a 2D/3D registration scheme and later related to the patient leg pose which is calculated using video-tracked AR markers. For the distal locking, surgeons use an augmented radiolucent drill in which its tip position is detected and tracked in real-time under video guidance.
Validation To evaluate the feasibility of our solution, we performed a preclinical study on dry bone phantom with the participation of four clinicians.
Results Participants achieved 100% success rate in the down-the beam positioning and 93% success rate in distal locking using only two X-ray images in 100 s. Conclusions We confirmed that intra-operative navigation using augmented reality provides an alternative way to perform distal locking in a safe and timely manner.

International Journal of Computer Assisted Radiology and Surgery, 2015.
Reprinted with permission from Springer.
DOI: https://doi.org/10.1007/s11548-015-1169-2

A Stereo Vision Approach for Cooperative Robotic Movement Therapy

Benjamin Busam, Marco Esposito, Simon Che'Rose, Nassir Navab, Benjamin Frisch



Overview of the proposed system. The impaired arm is fixated to the robotic arm via a soft textile support. The optical tracking system held by the robot follows the position of circular markers attached to a soft sleeve worn on the healthy limb. The robot keeps the relative position of the two arms constant, allowing the user to control the robot movement intuitively.

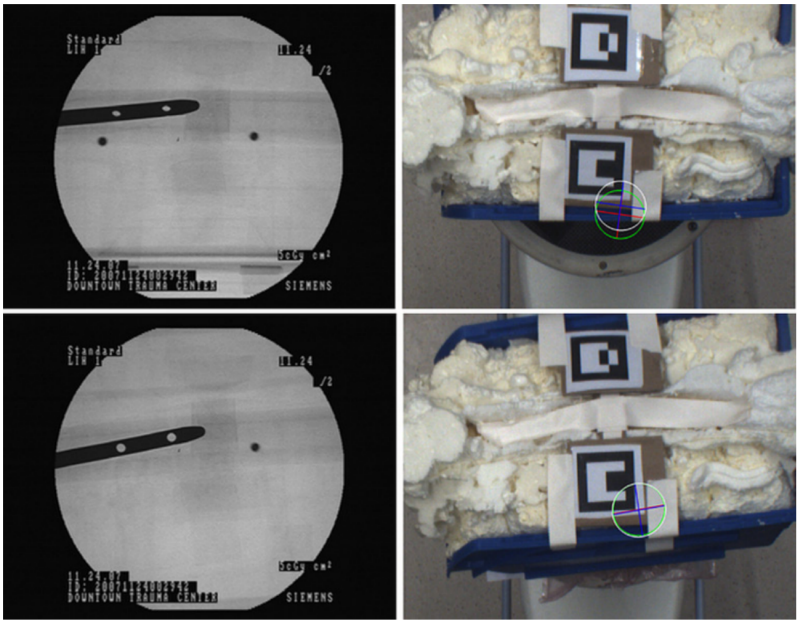
Movement therapy is an integrating part of stroke rehabilitation. The positive influence of intensive, repetitive motion training and the importance of active patient participation trigger the development of cooperative robotic assistants. We suggest a device for the re-education of upper limb movements in hemiparetic patients where a light-weight robotic arm that supports the deficient arm is equipped with a stereoscopic camera system. It follows the movements of the healthy arm that wears a sleeve equipped with flat round reflective markers detected by the cameras. We introduce an advanced robust and real-time algorithm to provide the tracking information. It performs a sparse marker based point cloud registration based on subpixel precision contour fits to enable high accuracy pose estimates while being capable of online model adjustments. The update rate of the tracking is 9 ms and the precision of the system is measured to be 0.5 mm. Tests with healthy subjects show that the system is able to accurately reproduce the movement of the healthy arm on an impaired arm.

IEEE International Conference on Computer Vision Workshop (ICCVW), 2015

DOI: https://doi.org/10.1109/ICCVW.2015.74

The 'Augmented' Circles: A Video-Guided Solution for the Down-the-Beam Positioning of IM Nail Holes

Roberto Londei, Marco Esposito, Benoit Diotte, Simon Weidert, Ekkehard Euler, Peter Thaller, Nassir Navab, Pascal Fallavollita



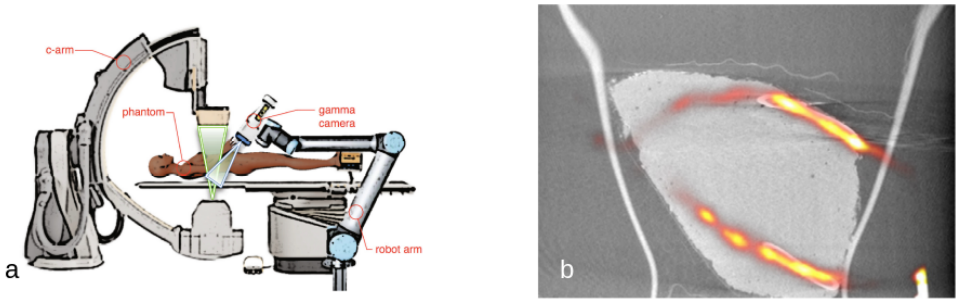
The initial X-ray image can be seen in the top left corner. From this single X-ray image the full 6 Degrees of Freedom pose of the intramedullary nail is retrieved, and an Augmented Reality guidance is provided to the user in the form of two circles (top right). After moving the X-ray machine such that the two circles coincide (bottom right), the Down-the-Beam position has been reached (bottom left) and the screw placing can begin.

Intramedullary nailing is the surgical procedure mostly used in fracture reduction of the tibial and femoral shafts. Following successful insertion of the nail into the medullary canal, it must be fixed by inserting screws through its proximal and distal locking holes. Prior to distal locking of the nail, surgeons must position the C-arm device and patient leg in such a way that the nail holes appear as circles in the X-ray image. This is considered a 'trial and error' process, is time consuming and requires many X-ray shots. We propose an augmented reality application that visually depicts to the surgeon two 'augmented' circles, their centers lying on the axis of the nail hole, making it visible in space. After an initial X-ray image acquisition, real-time video guidance allows the surgeon to superimpose the 'augmented' circles by moving the patient leg; the result being nail holes appearing as circles. Our nail pose recovery was evaluated on 1000 random trials and we consistently recovered the nail angulation within $2.76 \pm 1.66°$. Lastly, in a preclinical experiment involving 7 clinicians, we demonstrated that in over 95% of the trials, the nail hole appeared as a circle using an initial X-ray image.

Information Processing in Computer-Assisted Interventions. IPCAI 2014.
Reprinted with permission from Springer.
DOI: https://doi.org/10.1007/978-3-319-07521-1_11

Towards personalized interventional SPECT-CT imaging

José Gardiazabal, Marco Esposito, Philipp Matthies, Aslı Okur, Jakob Vogel, Silvan Kraft, Benjamin Frisch, Tobias Lasser, Nassir Navab



(a) System overview: a robot holding a portable gamma camera is introduced to the operating room and co-calibrated with a standard Cone-Beam CT device. This allows to acquire co-registered CT and SPECT volumes, which can be combined together into a single multi-modal image (b).

The development of modern robotics and compact imaging detectors allows the transfer of diagnostic imaging modalities to the operating room, supporting surgeons to perform faster and safer procedures. An intervention that currently suffers from a lack of interventional imaging is radioembolization, a treatment for hepatic carcinoma. Currently, this procedure requires moving the patient from an angiography suite for preliminary catheterization and injection to a whole-body SPECT/CT for leakage detection, necessitating a second catheterization back in the angiography suite for the actual radioembolization. We propose an imaging setup that simplifies this procedure using a robotic approach to directly acquire an interventional SPECT/CT in the angiography suite. Using C-arm CT and a co-calibrated gamma camera mounted on a robotic arm, a personalized trajectory of the gamma camera is generated from the C-arm CT, enabling an interventional SPECT reconstruction that is inherently co-registered to the C-arm CT. In this work we demonstrate the feasibility of this personalized interventional SPECT/CT imaging approach in a liver phantom study.

International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2014.

DOI: https://doi.org/10.1007/978-3-319-10404-1_63

C

Software created or contributed to in the Scope of this Thesis

During the work reported in this Thesis, the IFL laboratory underwent a process of reorganization of existing software and of development of new shared components. Some of them were made publicly available on a common GitHub page [1]. In this section we report the projects with major contribution by the author of this thesis.

iiwa_stack

iiwa_stack is a software package for the native integration of the KUKA LBR iiwa robots with the ROS ecosystem. Particular effort was spent in order to seamlessly fit the standard ROS interfaces to the KUKA Sunrise software and to provide extensive and detailed documentation. The development began jointly with Salvatore Virga, who was coauthor of many publications listed in Appendices A and B and overtook the role of maintainer of the library.
The native Sunrise development environment of the iiwa requires to write a Robotic Application in the Java programming language, which must then be deployed on the robot controller with a proprietary tool. The application runs in a non-realtime environment, but periodically communicates with the real-time Operating System. This architecture allows to maintain the safety requirements imposed by the industrial certifications for human collaboration regardless of the behavior of the Robotic Application.
Our library consists of a Robotic Application including rosjava nodes that expose a standard ROS interface (topics and services) over the Ethernet port of the controller. This allows to control the robot's SmartServo and DirectServo controllers (which provide advanced interpolation and safety features) over the network in real time. Alternatively, MoveIt integration is offered to enable high-level planning and obstacle avoidance through an RGBD camera.
iiwa_stack enjoys growing popularity, also among worldwide renowned robotic challenges, and a highly active community.

easy_handeye

Chapters 5 and 6 provide extensive sections about calibration. Various instances of hand-eye calibration between a robot and a tracking sensor contributes are listed within. Traditionally, hand-eye calibration is performed with programs or scripts designed ad-hoc for the particular setup and hardware, and output the result in their own format; the content must often be copy-pasted into a file, or hardcoded. Such approach is suboptimal to say the least, in particular when the same calibration must be repeated frequently.
This encouraged the author to develop a software library for the acquisition of data sam-

[1] https://github.com/IFL-CAMP

ples, computation of the calibration and easy usage of the result with the minimal number of manual steps involved. Furthermore, this framework makes use of ROS interfaces and infrastructure to be completely hardware agnostic and trivial to integrate.
The core component is a server which can be configured through a *.yaml* file to sample the transformations between specified tf reference frames upon request. The collected data is passed to a library implementing a hand-eye calibration method (currently only the implementation of the Tsai-Lenz algorithm from the ViSP library is supported, but extension would be straightforward). The result is stored in a file with a given ID. A second program can be started by the user's roslaunch files with the same ID as a parameter, and its purpose is to publish the calibration result on the tf subsystem. In this way all the running ROS processes are informed of the current calibration, eliminating chances of error: repeating the calibration will overwrite the file, thus leading to the updated value being automatically loaded at the next system restart. Using different IDs, multiple calibrations can be used concurrently.
Finally, a script was developed to automatically move the robot about the starting position about all axes, in a fashion apt to obtain a robust calibration. Intuitive GUIs were developed for the sampling and robot movement scripts.
easy_handeye is adopted by a growing number of users, in particular to use RGBD cameras for object avoidance, and was presented at ROSCON 2017.

tf_bag

The tf subsystem greatly simplifies the task of managing the knowledge of the current geometric configuration of the system. Traditional approaches rely on matrices being passed around in code as arguments, or over network connections. It is also common to coalesce whole chains of transformations within a single matrix for optimization purposes. Such methods ar error prone, verbose and complex for large systems.
Instead, ROS processes can be trivially integrated with tf to obtain knowledge of the transformation between two arbitrary reference frames at any point in time, or to share such knowledge with other modules or processes. tf identifies each reference frame with a string ID; the transformation between to such frames A and B at a certain point in time t can be queried from the system, or broadcasted into it in the form of ROS messages on the *tf* topic. Each process registered to the tf subsystem holds a buffer of all the messages published on this topic, so that upon a query for the transformation between two frames A and D the whole chain can be composed (e.g. A to B, B to C and so on). The library also supports queries for the relative position of two frames in the past (within the length of the buffer).
Since tf is implemented on top of a standard ROS topic, its data can be recorded via rosbag as with any other type of message. Anyway, extracting tf data from a rosbag file is not an easy task. The rosbag interface allows to linearly scan the recorded stream of messages; the user must then scan the whole file until the target time t, while keeping note of the last known values of all the reference frames joining the two frames of interest.
tf_bag is a library that offers an interface similar to what tf makes available at runtime, making it easy to analyze tf data contained in a rosbag (for example in scripts). More functionality for simple data analysis was added, and the operations were highly optimized for common use cases.

D Implementing the Tracking Regularization Algorithm

In order to implement the proposed scheme in Alg. 1, we have to implement the functions `compLambda`, `proxData`, `proxR1`, and `proxR2` for $SE(3)$. In the following parts, we will explain the necessary mathematical background for implementing these functions.

D.1 Second Order Total Variation Regularizer

As proposed by Bacák *et al.* [2], it is possible to generalize the idea of Eq. 6.15 to higher order differences:

$$R_2(\mathbf{x}) = \sum_{i=2}^{k-1} (h \circ d)(\mathbf{x}_i, \mathbf{m}_{i-1,i+1}), \tag{D.1}$$

where

$$\mathbf{m}_{i-1,i+1} = \exp_{\mathbf{x}_{i-1}}(\frac{1}{2}\log_{\mathbf{x}_{i-1}}(\mathbf{x}_{i+1})) \tag{D.2}$$

denotes the geodesic average of $\mathbf{x}_{i-1}$ and $\mathbf{x}_{i+1}$. As $d(\mathbf{x}_i, \mathbf{m}_{i-1,i+1})$ can be considered as a manifold-valued, second-order central difference, R_2 can be interpreted as a second-order approximation of the classical Tikhonov regularizer, in case of $h(s) = s^2/2$, or the second-order total variation, in case of $h(s) = s$, respectively. Likewise, one can also consider the Huber-type function in Eq. 6.14. Again, we will use the abbreviations TV ($\ell_1 - case$), TKHV ($\ell_2 - case$), and HUBER to denote the different regularization types.

D.2 Implementing the Proposed Algorithm

In order to implement the proposed scheme in Alg. 1, we have to implement the functions `compLambda`, `proxData`, `proxR1`, and `proxR2` for $SE(3)$. In the following parts, we will explain the necessary mathematical background for implementing these functions.

D.2.1 Exponential and Inverse Exponential Map

In order to realize the aforementioned proximal mappings, we need to implement the exponential and inverse exponential map for $SE(3)$. For the exponential mapping, we consider a

vector $v \in \mathfrak{se}(3)$ (sitting in the identity and representing a tangential vector by left translation) given by

$$v = \begin{pmatrix} \omega_v & t_v \\ 0 & 0 \end{pmatrix}, \tag{D.3}$$

as in Eq. 6.7, where $t_v \in \mathbb{R}^3$ is the $\mathfrak{so}_3$ part of v given by Eq. 6.8. The geodesics induced by the Riemannian metric Eq. 6.9 on $SE(3)$ are precisely the geodesics in the product manifold $SO(3) \times \mathbb{R}^3$, where $SO(3)$ is equipped with its bi-invariant metric. Therefore, for the pose $\mathbf{p}$,

$$\mathbf{p} = \begin{pmatrix} R_{\mathbf{p}} & t_{\mathbf{p}} \\ 0 & 1 \end{pmatrix}, \tag{D.4}$$

we have

$$\exp_{\mathbf{p}}(v) = \begin{pmatrix} \exp(\omega_v) R_{\mathbf{p}} & t_{\mathbf{p}} + t_v \\ 0 & 1 \end{pmatrix}. \tag{D.5}$$

Here, $\exp(\omega_v)$ denotes the matrix exponential function of $\omega_v \in \mathfrak{so}_3$. For the matrix operations needed above there are closed form expression available: To compute the matrix exponential of the skew-symmetric matrix ω_v, we use Rodrigues formula as explained in [123]; see also [199]. Then, for the inverse of the Riemannian exponential mapping in $SE(3)$ denoted by $\log$ with the left-invariant Riemann metric Eq. 6.9, we get for poses p, q that

$$\log_{\mathbf{p}}(\mathbf{q}) = \begin{pmatrix} \log(R_{\mathbf{q}} R_{\mathbf{p}}^t) & t_{\mathbf{q}} - t_{\mathbf{p}} \\ 0 & 0 \end{pmatrix}, \tag{D.6}$$

where $\log(R_{\mathbf{q}} R_{\mathbf{p}}^t)$ denotes the principal logarithm of the matrix $R_{\mathbf{q}} R_{\mathbf{p}}^t$ (which may be viewed as component-wise principal logarithm on the eigenvalues). Concerning the computation of the principal matrix logarithm of the rotation matrix $R_{\mathbf{q}} R_{\mathbf{p}^t}$, we again refer to the above mentioned references. We note that $\log_{\mathbf{p}}(\mathbf{q})$ is an element of the Lie algebra $\mathfrak{so}_3$ representing a tangent vector at $\mathbf{p}$ (up to left translation). We further note that the distance between poses $\mathbf{p}$ and $\mathbf{q}$ is explicitly given by

$$d(\mathbf{p}, \mathbf{q}) = \| \log(R_{\mathbf{q}} R_{\mathbf{p}^t}) \|_F + \| t_{\mathbf{p}} - t_{\mathbf{q}} \|, \tag{D.7}$$

where $\| \log(R_{\mathbf{q}} R_{\mathbf{p}^t}) \|_F$ denotes the Frobenius norm of $\log(R_{\mathbf{q}} R_{\mathbf{p}^t})$, and $\| t_{\mathbf{p}} - t_{\mathbf{q}} \|$ denotes the euclidean norm of $t_{\mathbf{p}} - t_{\mathbf{q}}$.

D.2.2 Computation of the Proximal Mappings

As a reference for the following derivations of the proximal mappings of the data and the first order difference terms, we refer to [199]. The proximal mapping for the data term atom $D_i(\mathbf{x}_i) = h \circ d(\mathbf{x}_i, \mathbf{p}_i)$, for a given data point $\mathbf{p}_i$ is given by

$$\operatorname{prox}_{\lambda D_i}(\mathbf{x}_i) = \exp_{\mathbf{x}_i} \left(t \ \log_{\mathbf{x}_i} \mathbf{p}_i \right) \tag{D.8}$$

where the parameter t is chosen, depending on the kind of data term used as, $t = \lambda/(1+\lambda)$ for the ℓ^2-type data term, by

$$t = \begin{cases} \lambda/d(\mathbf{x}_i, \mathbf{p}_i), & \text{if } \lambda < d(\mathbf{x}_i, \mathbf{p}_i), \\ d(\mathbf{x}_i, \mathbf{p}_i), & \text{else.} \end{cases} \tag{D.9}$$

for the ℓ^1 type data term, and by

$$t = \begin{cases} \frac{2\lambda}{1+2\lambda}, & \text{if } d(\mathbf{x}_i, \mathbf{p}_i) < \frac{\omega(1+2\lambda)}{\sqrt{2}}, \\ \min\left(d(\mathbf{x}_i, \mathbf{p}_i), \sqrt{2}\lambda\right)/d(\mathbf{x}_i, \mathbf{p}_i), & \text{otherwise,} \end{cases} \tag{D.10}$$

for the Huber type data term. The proximal mapping for the TV, the Huber and the analogue of the classical Tichonov regularizer atoms $R_{1,i}(\mathbf{x}_i, \mathbf{x}_{i+1}) = h \circ d(\mathbf{x}_i, \mathbf{x}_{i+1})$, are given by

$$\begin{aligned} (\operatorname{prox}_{\lambda R_{1,i}}\mathbf{x})_i &= \exp_{\mathbf{x}_i}\left(t \ \log_{\mathbf{x}_i}\mathbf{x}_{i+1}\right), \\ (\operatorname{prox}_{\lambda R_{1,i}}\mathbf{x})_{i+1} &= \exp_{\mathbf{x}_{i+1}}\left(t \ \log_{\mathbf{x}_{i+1}}\mathbf{x}_i\right), \end{aligned}$$

where $t = \lambda/(1+2\lambda)$ for the ℓ_2-type weighting, where

$$t = \begin{cases} \lambda, & \text{if } \lambda < d(\mathbf{x}_i, \mathbf{x}_{i+1})/2, \\ d(\mathbf{x}_i, \mathbf{p}_i)/2, & \text{else,} \end{cases} \tag{D.11}$$

for the TV situation, and by

$$t = \begin{cases} \frac{2\lambda}{1+4\lambda}, & \text{if } d(\mathbf{x}_i, \mathbf{x}_{i+1}) < \frac{1+4\lambda}{\sqrt{2}}, \\ \min\left(d(\mathbf{x}_i, \mathbf{x}_{i+1})/2, \sqrt{2}\lambda\right)/d(\mathbf{x}_i, \mathbf{p}_i), & \text{otherwise.} \end{cases} \tag{D.12}$$

in case of Huber regularization. In contrast to the proximal mappings of the data and the first order difference terms, the proximal mappings for the second order TV type terms R_2 do not have known closed form expressions. Instead, we use a subgradient descent scheme to compute the proximal mappings, which are the solutions of the minimization problems

$$\begin{aligned} &\operatorname{prox}_{\lambda R_{2,i}}(\mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1}) \\ &= \operatorname*{argmin}_{(\mathbf{x}'_{i-1}, \mathbf{x}'_i, \mathbf{x}'_{i+1})} \sum_{j=i-1}^{i+1} \frac{1}{2} d(\mathbf{x}_j, \mathbf{x}'_j)^2 + R_{2,i}(\mathbf{x}'_{i-1}, \mathbf{x}'_i, \mathbf{x}'_{i+1}) \end{aligned}$$

approximately. The corresponding gradients can be computed rather explicitly using Jacobi fields. Details can be found in [2].

D.2.3 Choice of Step Size

During the iteration of Alg. 1, the step size parameter λ_r of the proximal mappings is successively decreased. In this way, the penalty for deviation from the previous iterate is successively increased. It is chosen in a way such that the sequence λ_r is square-summable but not

summable. This is moderate enough not to enforce convergence by step size decay. Concretely, we use the sequence

$$\lambda_r = 0.25r^{-0.95} \tag{D.13}$$

to realize this decay in the algorithmic realization.

Bibliography

[1] A. Agarwal, C. Nasa, and S. Bandyopadhyay. “Dynamic singularity avoidance for parallel manipulators using a task-priority based control scheme”. In: *Mechanism and Machine Theory* 96 (Feb. 2016). 00018, pp. 107–126. DOI: 10.1016/j.mechmachtheory.2015.07.013 (cit. on p. 68).

[2] M. Bačák, R. Bergmann, G. Steidl, and A. Weinmann. “A Second Order Nonsmooth Variational Model for Restoring Manifold-Valued Images”. In: *SIAM Journal on Scientific Computing* 38.1 (Jan. 2016). 00000, A567–A597. DOI: 10.1137/15M101988X (cit. on pp. 101, 103, 139, 141).

[3] P. A. Bandettini, A. Jesmanowicz, E. C. Wong, and J. S. Hyde. “Processing strategies for time-course data sets in functional mri of the human brain”. In: *Magnetic Resonance in Medicine* 30.2 (Aug. 1993). 02098, pp. 161–173. DOI: 10.1002/mrm.1910300204 (cit. on pp. 24, 25).

[4] W. L. Bargar, A. Bauer, and M. Börner. “Primary and Revision Total Hip Replacement Using the Robodoc System”. In: *Clinical Orthopaedics and Related Research* 354 (Sept. 1998). 00419, pp. 82–91. DOI: 10.1097/00003086-199809000-00011 (cit. on p. 48).

[5] C. Barry, C. Allott, N. John, P. Mellor, P. Arundel, D. Thomson, and J. Waterton. “Three-dimensional freehand ultrasound: Image reconstruction and volume analysis”. In: *Ultrasound in Medicine & Biology* 23.8 (Jan. 1997), pp. 1209–1224. DOI: 10.1016/S0301-5629(97)00123-3 (cit. on p. 98).

[6] A. Bartoli, T. Collins, N. Bourdel, and M. Canis. “Computer assisted Minimally Invasive Surgery: Is medical Computer Vision the answer to improving laparosurgery?” In: *Medical Hypotheses* 79.6 (Dec. 2012). 00011, pp. 858–863. DOI: 10.1016/j.mehy.2012.09.007 (cit. on p. 36).

[7] R. A. Beasley. “Medical Robots: Current Systems and Research Directions”. In: *Journal of Robotics* 2012 (2012). 00148, pp. 1–14. DOI: 10.1155/2012/401613 (cit. on p. 47).

[8] C. Belta and V. Kumar. “Euclidean metrics for motion generation on *SE* (3)”. In: *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 216.1 (Jan. 2002). 00000, pp. 47–60. DOI: 10.1243/0954406021524909 (cit. on p. 101).

[9] J. van den Berg, S. Miller, D. Duckworth, H. Hu, A. Wan, Xiao-Yu Fu, K. Goldberg, and P. Abbeel. “Superhuman performance of surgical tasks by robots using iterative learning from human-guided demonstrations”. In: *2010 IEEE International Conference on Robotics and Automation*. 2010 IEEE International Conference on Robotics and Automation (ICRA 2010). 00130. Anchorage, AK: IEEE, May 2010, pp. 2074–2081. DOI: 10.1109/ROBOT.2010.5509621 (cit. on p. 53).

[10] R. Berguer and W. Smith. “An Ergonomic Comparison of Robotic and Laparoscopic Technique: The Influence of Surgeon Experience and Task Complexity”. In: *Journal of Surgical Research* 134.1 (July 2006). 00109, pp. 87–92. DOI: 10.1016/j.jss.2005.10.003 (cit. on pp. 34, 50).

[11] S. Bernhardt, S. A. Nicolau, L. Soler, and C. Doignon. “The status of augmented reality in laparoscopic surgery as of 2016”. In: *Medical Image Analysis* 37 (Apr. 2017). 00042, pp. 66–90. DOI: 10.1016/j.media.2017.01.007 (cit. on pp. 36, 37, 42).

[12] P. J. Besl and N. D. McKay. “Method for registration of 3-D shapes”. In: Robotics - DL tentative. Ed. by P. S. Schenker. 16353. Boston, MA, Apr. 30, 1992, pp. 586–606. DOI: 10.1117/12.57955 (cit. on p. 113).

[13] M. S. Bhamla, B. Benson, C. Chai, G. Katsikis, A. Johri, and M. Prakash. “Hand-powered ultralow-cost paper centrifuge”. In: *Nature Biomedical Engineering* 1.1 (Jan. 10, 2017). 00044, p. 0009. DOI: 10.1038/s41551-016-0009 (cit. on p. 121).

[14] W. Birkfellner, F. Watzinger, F. Wanschitz, G. Enislidis, C. Kollmann, D. Rafolt, R. Nowotny, R. Ewers, and H. Bergmann. “Systematic distortions in magnetic position digitizers”. In: *Medical Physics* 25.11 (Nov. 1998). 00143, pp. 2242–2248. DOI: 10.1118/1.598425 (cit. on p. 95).

[15] J. E. Boland, L. W. Wang, B. J. Love, M. Christofi, and D. W. Muller. “Impact of New-generation Hybrid Imaging Technology on Radiation Dose during Percutaneous Coronary Interventions and Trans-femoral Aortic Valve Implantations: A comparison with conventional flat-plate angiography”. In: *Heart, Lung and Circulation* 25.7 (July 2016). 00004, pp. 668–675. DOI: 10.1016/j.hlc.2015.08.023 (cit. on p. 49).

[16] M. Brubaker, M. Salzmann, and R. Urtasun. “A Family of MCMC Methods on Implicitly Defined Manifolds”. In: *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*. Ed. by N. D. Lawrence and M. Girolami. Vol. 22. Proceedings of Machine Learning Research. 00042. La Palma, Canary Islands: PMLR, Apr. 21, 2012, pp. 161–172 (cit. on p. 103).

[17] J. N. Bruneton, E. Caramella, M. Héry, D. Aubanel, J. J. Manzino, and J. L. Picard. “Axillary lymph node metastases in breast cancer: preoperative detection with US.” In: *Radiology* 158.2 (Feb. 1986). 00226, pp. 325–326. DOI: 10.1148/radiology.158.2.3510440 (cit. on p. 59).

[18] S. L. Bugby, J. E. Lees, and A. C. Perkins. “Hybrid intraoperative imaging techniques in radioguided surgery: present clinical applications and future outlook”. In: *Clinical and Translational Imaging* 5.4 (Aug. 2017). 00007, pp. 323–341. DOI: 10.1007/s40336-017-0235-x (cit. on p. 59).

[19] J. Burgner-Kahrs, D. C. Rucker, and H. Choset. “Continuum Robots for Medical Applications: A Survey”. In: *IEEE Transactions on Robotics* 31.6 (Dec. 2015). 00173, pp. 1261–1280. DOI: 10.1109/TRO.2015.2489500 (cit. on p. 53).

[20] B. Busam, T. Birdal, and N. Navab. “Camera Pose Filtering with Local Regression Geodesics on the Riemannian Manifold of Dual Quaternions”. In: *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*. 2017 IEEE International Conference on Computer Vision Workshop (ICCVW). 00005. Venice: IEEE, Oct. 2017, pp. 2436–2445. DOI: 10.1109/ICCVW.2017.287 (cit. on p. 101).

[21] B. Busam, M. Esposito, S. Che’Rose, N. Navab, and B. Frisch. “A Stereo Vision Approach for Cooperative Robotic Movement Therapy”. In: *IEEE International Conference on Computer Vision Workshop (ICCVW)*. 2015 (cit. on pp. 73, 76, 126).

[22] B. Busam, M. Esposito, B. Frisch, and N. Navab. “Quaternionic Upsampling: Hyperspherical Techniques for 6 DoF Pose Tracking”. In: *IEEE International Conference on 3DVision*. 2016 (cit. on p. 125).

[23] G. B. Cadière, J. Himpens, O. Germay, R. Izizaw, M. Degueldre, J. Vandromme, E. Capelluto, and J. Bruyns. “Feasibility of robotic laparoscopic surgery: 146 cases”. In: *World J Surg* 25.11 (Nov. 2001), pp. 1467–1477 (cit. on p. 11).

[24] C. M. Carswell, D. Clarke, and W. B. Seales. “Assessing Mental Workload During Laparoscopic Surgery”. In: *Surgical Innovation* 12.1 (Mar. 2005). 00127, pp. 80–90. DOI: 10.1177/155335060501200112 (cit. on pp. 34, 42).

[25] M. Chafkin. “Robots Could Replace Surgeons in the Battle Against Cancer”. In: *Bloomberg Businessweek* (Mar. 23, 2018). 00000 (cit. on p. 120).

[26] S. R. Cherry. “Multimodality Imaging: Beyond PET/CT and SPECT/CT”. In: *Seminars in Nuclear Medicine* 39.5 (Sept. 2009). 00231, pp. 348–353. DOI: 10.1053/j.semnuclmed.2009.03.001 (cit. on pp. 7, 31).

[27] H. M. Choset, ed. *Principles of robot motion: theory, algorithms, and implementation*. Intelligent robotics and autonomous agents. 02826 OCLC: ocm54460979. Cambridge, Mass: MIT Press, 2005. 603 pp. (cit. on p. 67).

[28] F. Corcione, C. Esposito, D. Cuccurullo, A. Settembre, N. Miranda, F. Amato, F. Pirozzi, and P. Caiazzo. “Advantages and limits of robot-assisted laparoscopic surgery: preliminary experience”. In: *Surgical Endoscopy* 19.1 (Jan. 2005). 00247, pp. 117–119. DOI: 10.1007/s00464-004-9004-9 (cit. on pp. 7, 11, 54).

[29] V. Cordemans, L. Kaminski, X. Banse, B. G. Francq, and O. Cartiaux. “Accuracy of a new intra-operative cone beam CT imaging technique (Artis zeego II) compared to postoperative CT scan for assessment of pedicle screws placement and breaches detection”. In: *European Spine Journal* 26.11 (Nov. 2017). 00004, pp. 2906–2916. DOI: 10.1007/s00586-017-5139-y (cit. on p. 49).

[30] P. Corke and M. Good. “Dynamic effects in high-performance visual servoing”. In: *Proceedings 1992 IEEE International Conference on Robotics and Automation*. 1992 IEEE International Conference on Robotics and Automation. 00066. Nice, France: IEEE Comput. Soc. Press, 1992, pp. 1838–1843. DOI: 10.1109/ROBOT.1992.219960 (cit. on p. 74).

[31] K. Cox, A. Sever, S. Jones, J. Weeks, P. Mills, H. Devalia, D. Fish, and P. Jones. “Validation of a technique using microbubbles and contrast enhanced ultrasound (CEUS) to biopsy sentinel lymph nodes (SLN) in pre-operative breast cancer patients with a normal grey-scale axillary ultrasound”. In: *European Journal of Surgical Oncology (EJSO)* 39.7 (July 2013), pp. 760–765. DOI: 10.1016/j.ejso.2013.03.026 (cit. on p. 61).

[32] J. J. Craig. *Introduction to robotics: mechanics and control*. 3rd ed. 10756. Upper Saddle River, N.J: Pearson/Prentice Hall, 2005. 400 pp. (cit. on p. 66).

[33] J. Dang, B. Frisch, P. Lasaygues, D. Zhang, S. Tavernier, N. Felix, P. Lecoq, E. Auffray, J. Varela, S. Mensah, and M. Wan. “Development of an Anthropomorphic Breast Phantom for Combined PET, B-Mode Ultrasound and Elastographic Imaging”. In: *IEEE Transactions on Nuclear Science* 58.3 (June 2011). 00021, pp. 660–667. DOI: 10.1109/TNS.2011.2105279 (cit. on p. 82).

[34] R. O. Deaner and M. L. Platt. “Reflexive Social Attention in Monkeys and Humans”. In: *Current Biology* 13.18 (Sept. 2003). 00159, pp. 1609–1613. DOI: 10.1016/j.cub.2003.08.025 (cit. on p. 52).

[35] J. Denavit and R. S. Hartenberg. “A Kinematic Notation for Lower-Pair Mechanisms Based on Matrices”. In: *Trans. ASME, J. Appl. Mech.* 22.2 (1965). 00000, pp. 215–221 (cit. on p. 67).

[36] G. Dogangil, B. L. Davies, and F. Rodriguez y Baena. “A review of medical robotics for minimally invasive soft tissue surgery”. In: *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* 224.5 (May 2010), pp. 653–679. DOI: 10.1243/09544119JEIM591 (cit. on pp. 38, 45, 46).

[37] A. D. Dragan, S. Bauman, J. Forlizzi, and S. S. Srinivasa. “Effects of Robot Motion on Human-Robot Collaboration”. In: *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI ’15*. the Tenth Annual ACM/IEEE International Conference. 00078. Portland, Oregon, USA: ACM Press, 2015, pp. 51–58. DOI: 10.1145/2696454.2696473 (cit. on p. 52).

[38] J. P. Ehlers, S. K. Srivastava, D. Feiler, A. I. Noonan, A. M. Rollins, and Y. K. Tao. “Integrative Advances for OCT-Guided Ophthalmic Surgery and Intraoperative OCT: Microscope Integration, Surgical Instrumentation, and Heads-Up Display Surgeon Feedback”. In: *PLoS ONE* 9.8 (Aug. 20, 2014). Ed. by A. S. Lewin, e105224. DOI: 10.1371/journal.pone.0105224 (cit. on p. 29).

[39] A. Elsherbiny, S. Koller, N. Kohn, D. Ostler, A. Schneider, T. Vogel, D. Wilhelm, H. Friess, H. Feussner, and M. Kranzfelder. “Evaluation of Eye-Tracking vs Color-code Tracking for Robotic Camera Assistance in Minimally Invasive Surgery”. In: (2017) (cit. on p. 53).

[40] M. Esposito, B. Busam, C. Hennersperger, J. Rackerseder, A. Lu, N. Navab, and B. Frisch. “Cooperative Robotic Gamma Imaging: Enhancing US-guided Needle Biopsy”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Ed. by N. Navab, J. Hornegger, W. M. Wells, and A. Frangi. Vol. 9350. 00009. Cham: Springer International Publishing, 2015, pp. 611–618. DOI: 10.1007/978-3-319-24571-3_73 (cit. on pp. 57, 81, 126).

[41] M. Esposito, B. Busam, C. Hennersperger, J. Rackerseder, N. Navab, and B. Frisch. “Multimodal US–gamma imaging using collaborative robotics for cancer staging biopsies”. In: *International Journal of Computer Assisted Radiology and Surgery* 11.9 (Sept. 2016). 00004, pp. 1561–1571. DOI: 10.1007/s11548-016-1464-6 (cit. on pp. 57, 125).

[42] M. Esposito, C. Hennersperger, R. Gobl, L. Demaret, M. Storath, N. Navab, M. Baust, and A. Weinmann. “Total Variation Regularization of Pose Signals With an Application to 3D Freehand Ultrasound”. In: *IEEE Trans. Med. Imaging* 38.10 (Oct. 2019), pp. 2245–2258. DOI: 10.1109/TMI.2019.2898480 (cit. on pp. 90, 125).

[43] M. P. Esposito, P. Ilbeigi, M. Ahmed, and V. Lanteri. “Use of fourth arm in da Vinci robot-assisted extraperitoneal laparoscopic prostatectomy: Novel technique”. In: *Urology* 66.3 (Sept. 2005). 00053, pp. 649–652. DOI: 10.1016/j.urology.2005.03.061 (cit. on p. 11).

[44] J. Esteban, W. Simson, S. Requena Witzig, A. Rienmüller, S. Virga, B. Frisch, O. Zettinig, D. Sakara, Y.-M. Ryang, N. Navab, and C. Hennersperger. “Robotic ultrasound-guided facet joint insertion”. In: *International Journal of Computer Assisted Radiology and Surgery* 13.6 (June 2018). 00000, pp. 895–904. DOI: 10.1007/s11548-018-1759-x (cit. on p. 53).

[45] C. Fabbri. “Endoscopic ultrasound-guided treatments: Are we getting evidence based - a systematic review”. In: *World Journal of Gastroenterology* 20.26 (2014). 00078, p. 8424. DOI: 10.3748/wjg.v20.i26.8424 (cit. on p. 26).

[46] A. F. Fercher, W. Drexler, C. K. Hitzenberger, and T. Lasser. “Optical coherence tomography - principles and applications”. In: *Reports on Progress in Physics* 66.2 (Feb. 1, 2003), pp. 239–303. DOI: 10.1088/0034-4885/66/2/204 (cit. on p. 29).

[47] H. Feußner and A. Park. “Surgery 4.0: the natural culmination of the industrial revolution?” In: *Innovative Surgical Sciences* 2.3 (Sept. 26, 2017). 00002, pp. 105–108. DOI: 10.1515/iss-2017-0036 (cit. on p. 50).

[48] H. Feußner and B. Rau. “Augmented Reality, Cyber-Physical Systems and Robotics: Nice to Have or a Program with Future?” In: *Visceral Medicine* 34.1 (2018). 00000, pp. 8–9. DOI: 10.1159/000487535 (cit. on p. 50).

[49] S. Finkelstein. “GM and the great automation solution”. In: *Business Strategy Review* 14.3 (Sept. 2003). 00008, pp. 18–24. DOI: 10.1111/1467-8616.00268 (cit. on p. 121).

[50] K. Fischer, H. M. Weigelin, and L. Bodenhagen. “Increasing trust in human–robot medical interactions: effects of transparency and adaptability”. In: *Paladyn, Journal of Behavioral Robotics* 9.1 (June 1, 2018). 00000, pp. 95–109. DOI: 10.1515/pjbr-2018-0007 (cit. on p. 52).

[51] 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 (Aug. 2014). 00128, pp. 1702–1725. DOI: 10.1109/TMI.2014.2321777 (cit. on pp. 94, 95, 97).

[52] B. Fuerst, J. Sprung, F. Pinto, B. Frisch, T. Wendler, H. Simon, L. Mengus, N. S. van den Berg, H. G. van der Poel, F. W. B. van Leeuwen, et al. “First robotic SPECT for minimally invasive sentinel lymph node mapping”. In: *IEEE transactions on medical imaging* 35.3 (2016), pp. 830–838 (cit. on p. 41).

[53] J. Gardiazabal, M. Esposito, P. Matthies, A. Okur, J. Vogel, S. Kraft, B. Frisch, T. Lasser, and N. Navab. “Towards Personalized Interventional SPECT-CT Imaging”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014*. Ed. by P. Golland, N. Hata, C. Barillot, J. Hornegger, and R. Howe. Cham: Springer International Publishing, 2014, pp. 504–511. DOI: 10.1007/978-3-319-10404-1_63 (cit. on pp. 49, 126).

[54] A. Gee, R. Prager, G. Treece, and L. Berman. “Engineering a freehand 3D ultrasound system”. In: *Pattern Recognition Letters* 24.4 (Feb. 2003). 00144, pp. 757–777. DOI: 10.1016/S0167-8655(02)00180-0 (cit. on pp. 26, 91).

[55] E. A. Geiser, L. G. Christie, D. A. Conetta, C. R. Conti, and G. S. Gossman. “A mechanical arm for spatial registration of two-dimensional echocardiographic sections”. In: *Catheterization and Cardiovascular Diagnosis* 8.1 (1982), pp. 89–101. DOI: 10.1002/ccd.1810080114 (cit. on p. 93).

[56] P. C. Giulianotti. “Robotics in General Surgery: Personal Experience in a Large Community Hospital”. In: *Archives of Surgery* 138.7 (July 1, 2003). 00749, p. 777. DOI: 10.1001/archsurg.138.7.777 (cit. on pp. 7, 50, 54).

[57] P. C. Giulianotti, F. Sbrana, F. M. Bianco, E. F. Elli, G. Shah, P. Addeo, G. Caravaglios, and A. Coratti. “Robot-assisted laparoscopic pancreatic surgery: single-surgeon experience”. In: *Surgical Endoscopy* 24.7 (July 2010). 00343, pp. 1646–1657. DOI: 10.1007/s00464-009-0825-4 (cit. on p. 50).

[58] M. Golatta, D. Franz, A. Harcos, H. Junkermann, G. Rauch, A. Scharf, F. Schuetz, C. Sohn, and J. Heil. “Interobserver reliability of automated breast volume scanner (ABVS) interpretation and agreement of ABVS findings with hand held breast ultrasound (HHUS), mammography and pathology results”. In: *European Journal of Radiology* 82.8 (Aug. 2013). 00072, e332–e336. DOI: 10.1016/j.ejrad.2013.03.005 (cit. on p. 26).

[59] I. Goldstein, E. Reece, G. Pilu, L. Bovicelli, and J. C. Hobbins. “Cerebellar measurements with ultrasonography in the evaluation of fetal growth and development”. In: *American Journal of Obstetrics and Gynecology* 156.5 (May 1987). 00166, pp. 1065–1069. DOI: 10.1016/0002-9378(87)90111-6 (cit. on p. 91).

[60] P. Gomes. “Surgical robotics: Reviewing the past, analysing the present, imagining the future”. In: *Robotics and Computer-Integrated Manufacturing* 27.2 (Apr. 2011). 00178, pp. 261–266. DOI: 10.1016/j.rcim.2010.06.009 (cit. on pp. 47, 50, 54).

[61] W. Grossman, ed. *Cardiac catheterization and angiography*. 3rd ed. 00716. Philadelphia: Lea & Febiger, 1986. 562 pp. (cit. on p. 23).

[62] S. Habert, J. Gardiazabal, P. Fallavollita, and N. Navab. “RGBDX: First Design and Experimental Validation of a Mirror-Based RGBD X-ray Imaging System”. In: *2015 IEEE International Symposium on Mixed and Augmented Reality*. 2015 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). Fukuoka, Japan: IEEE, Sept. 2015, pp. 13–18. DOI: 10.1109/ISMAR.2015.17 (cit. on p. 35).

[63] T. Haidegger, J. Sándor, and Z. Benyó. “Surgery in space: the future of robotic telesurgery”. In: *Surgical Endoscopy* 25.3 (Mar. 2011). 00112, pp. 681–690. DOI: 10.1007/s00464-010-1243-3 (cit. on p. 53).

[64] J. V. Hajnal, D. J. Hawkes, and D. L. G. Hill. *Medical image registration*. 02621 OCLC: 309875996. Boca Raton: CRC Press, 2001 (cit. on p. 30).

[65] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. 2nd ed. Cambridge, UK ; New York: Cambridge University Press, 2003. 655 pp. (cit. on pp. 64, 66).

[66] R. C. Harwell and R. L. Ferguson. “Physiologic tremor and microsurgery”. In: *Microsurgery* 4.3 (1983). 00118, pp. 187–192. DOI: 10.1002/micr.1920040310 (cit. on p. 49).

[67] K. Hasegawa, K. Noda, and Y. Sato. “Electronic endoscope system for shape measurement”. In: *Object recognition supported by user interaction for service robots*. 16th International Conference on Pattern Recognition. Vol. 2. Quebec City, Que., Canada: IEEE Comput. Soc, 2002, pp. 761–764. DOI: 10.1109/ICPR.2002.1048414 (cit. on p. 35).

[68] A. B. Haynes, T. G. Weiser, W. R. Berry, S. R. Lipsitz, A.-H. S. Breizat, E. P. Dellinger, T. Herbosa, S. Joseph, P. L. Kibatala, M. C. M. Lapitan, A. F. Merry, K. Moorthy, R. K. Reznick, B. Taylor, and A. A. Gawande. “A Surgical Safety Checklist to Reduce Morbidity and Mortality in a Global Population”. In: *New England Journal of Medicine* 360.5 (Jan. 29, 2009). 04242, pp. 491–499. DOI: 10.1056/NEJMsa0810119 (cit. on p. 121).

[69] J. Heikkila and O. Silven. “A four-step camera calibration procedure with implicit image correction”. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 02339. San Juan, Puerto Rico: IEEE Comput. Soc, 1997, pp. 1106–1112. DOI: 10.1109/CVPR.1997.609468 (cit. on p. 76).

[70] C. Hennersperger, M. Baust, D. Mateus, and N. Navab. “Computational sonography”. In: *International conference on medical image computing and computer-assisted intervention*. Springer, 2015, pp. 459–466 (cit. on p. 106).

[71] C. Hennersperger, B. Fuerst, S. Virga, O. Zettinig, B. Frisch, T. Neff, and N. Navab. “Towards MRI-Based Autonomous Robotic US Acquisitions: A First Feasibility Study”. In: *IEEE Transactions on Medical Imaging* 36.2 (Feb. 2017). 00021, pp. 538–548. DOI: 10.1109/TMI.2016.2620723 (cit. on p. 71).

[72] R. Hofstetter, M. Slomczykowski, M. Sati, and L.-P. Nolte. “Fluoroscopy as an Imaging Means for Computer-Assisted Surgical Navigation”. In: *Computer Aided Surgery* 4.2 (Jan. 1999). 00334, pp. 65–76. DOI: 10.3109/10929089909148161 (cit. on p. 38).

[73] M. Honl, O. Dierk, C. Gauck, V. Carrero, F. Lampe, S. Dries, M. Quante, K. Schwieger, E. Hille, and M. M. Morlock. “Comparison of robotic-assisted and manual implantation of a primary total hip replacement. A prospective study”. In: *J Bone Joint Surg Am* 85-A.8 (Aug. 2003). 00171, pp. 1470–1478 (cit. on p. 48).

[74] P. R. Hoskins, K. Martin, and A. Thrush, eds. *Diagnostic Ultrasound: Physics and Equipment*. 2nd ed. Cambridge: Cambridge University Press, 2010. DOI: 10.1017/CBO9780511750885 (cit. on p. 62).

[75] R. J. Housden, A. H. Gee, G. M. Treece, and R. W. Prager. “Sensorless Reconstruction of Unconstrained Freehand 3D Ultrasound Data”. In: *Ultrasound in Medicine & Biology* 33.3 (Mar. 2007). 00051, pp. 408–419. DOI: 10.1016/j.ultrasmedbio.2006.09.015 (cit. on pp. 89, 93, 94).

[76] J. Hsieh. *Computed tomography: principles, design, artifacts, and recent advances*. Third edition. 01661. Bellingham, Washington, USA: SPIE, 2015. 639 pp. (cit. on p. 23).

[77] P.-W. Hsu, G. M. Treece, R. W. Prager, N. E. Houghton, and A. H. Gee. “Comparison of Freehand 3-D Ultrasound Calibration Techniques Using a Stylus”. In: *Ultrasound in Medicine & Biology* 34.10 (Oct. 2008). 00044, pp. 1610–1621. DOI: 10.1016/j.ultrasmedbio.2008.02.015 (cit. on pp. 76, 77).

[78] P. J. Huber. “Robust Estimation of a Location Parameter”. In: *The Annals of Mathematical Statistics* 35.1 (Mar. 1964). 05131, pp. 73–101. DOI: 10.1214/aoms/1177703732 (cit. on p. 101).

[79] S. Hughes, T. D’Arcy, D. Maxwell, W. Chiu, A. Milner, J. Saunders, and R. Sheppard. “Volume estimation from multiplanar 2D ultrasound images using a remote electromagnetic position and orientation sensor”. In: *Ultrasound in Medicine & Biology* 22.5 (Jan. 1996). 00091, pp. 561–572. DOI: 10.1016/0301-5629(96)00022-1 (cit. on p. 93).

[80] L. Hull, S. Arora, R. Aggarwal, A. Darzi, C. Vincent, and N. Sevdalis. “The Impact of Nontechnical Skills on Technical Performance in Surgery: A Systematic Review”. In: *Journal of the American College of Surgeons* 214.2 (Feb. 2012). 00193, pp. 214–230. DOI: 10.1016/j.jamcollsurg.2011.10.016 (cit. on p. 42).

[81] F. A. Jolesz, ed. *Intraoperative Imaging and Image-Guided Therapy*. 00035. New York, NY: Springer New York, 2014. DOI: 10.1007/978-1-4614-7657-3 (cit. on pp. 33, 34, 37, 40).

[82] F. J. Joseph, A. van Oepen, and M. Friebe. “Breast sentinel lymph node biopsy with imaging towards minimally invasive surgery”. In: *Biomedical Engineering / Biomedizinische Technik* 62.6 (Nov. 27, 2017). 00001, pp. 547–555. DOI: 10.1515/bmt-2016-0164 (cit. on p. 59).

[83] J. H. Kaouk, R. Autorino, F. J. Kim, D. H. Han, S. W. Lee, S. Yinghao, J. A. Cadeddu, I. H. Derweesh, L. Richstone, L. Cindolo, A. Branco, F. Greco, M. Allaf, R. Sotelo, E. Liatsikos, J.-U. Stolzenburg, A. Rane, W. M. White, W. K. Han, G.-P. Haber, M. A. White, W. R. Molina, B. C. Jeong, J. Y. Lee, W. Linhui, S. Best, S. P. Stroup, S. Rais-Bahrami, L. Schips, P. Fornara, P. Pierorazio, C. Giedelman, J. W. Lee, R. J. Stein, and K. H. Rha. “Laparoendoscopic Single-site Surgery in Urology: Worldwide Multi-institutional Analysis of 1076 Cases”. In: *European Urology* 60.5 (Nov. 2011), pp. 998–1005. DOI: 10.1016/j.eururo.2011.06.002 (cit. on p. 47).

[84] L. R. Kavoussi, R. G. Moore, J. B. Adams, and A. W. Partin. “Comparison of Robotic Versus Human Laparoscopic Camera Control”. In: *The Journal of Urology* 154.6 (Dec. 1995). 00285, pp. 2134–2136. DOI: 10.1016/S0022-5347(01)66715-6 (cit. on p. 34).

[85] R. Khosla, A. W. McLean, and J. A. Smith. “Ultrasound-guided versus computed tomography-scan guided biopsy of pleural-based lung lesions”. In: *Lung India* 33.5 (Oct. 2016). 00002, pp. 487–492. DOI: 10.4103/0970-2113.188961 (cit. on p. 12).

[86] F. Kiessling, J. Huppert, and M. Palmowski. “Functional and Molecular Ultrasound Imaging: Concepts and Contrast Agents”. In: *Current Medicinal Chemistry* 16.5 (Feb. 1, 2009). 00111, pp. 627–642. DOI: 10.2174/092986709787458470 (cit. on p. 26).

[87] M. Klein, L. P. H. Andersen, M. Alamili, I. Gögenur, and J. Rosenberg. “Psychological and Physical Stress in Surgeons Operating in a Standard or Modern Operating Room:” in: *Surgical Laparoscopy, Endoscopy & Percutaneous Techniques* 20.4 (Aug. 2010), pp. 237–242. DOI: 10.1097/SLE.0b013e3181ed851d (cit. on p. 42).

[88] D. Krag, D. Weaver, J. Alex, and J. Fairbank. “Surgical resection and radiolocalization of the sentinel lymph node in breast cancer using a gamma probe”. In: *Surgical Oncology* 2.6 (Dec. 1993). 01983, pp. 335–340. DOI: 10.1016/0960-7404(93)90064-6 (cit. on p. 59).

[89] M. Kranzfelder, A. Schneider, A. Fiolka, S. Koller, D. Wilhelm, S. Reiser, A. Meining, and H. Feussner. “What Do We Really Need? Visions of an Ideal Human–Machine Interface for NOTES Mechatronic Support Systems From the View of Surgeons, Gastroenterologists, and Medical Engineers”. In: *Surgical Innovation* 22.4 (Aug. 2015). 00005, pp. 432–440. DOI: 10.1177/1553350614550720 (cit. on pp. 3, 53).

[90] A. Kumar and B. Asaf. “Robotic thoracic surgery: The state of the art”. In: *Journal of Minimal Access Surgery* 11.1 (2015). 00031, p. 60. DOI: 10.4103/0972-9941.147693 (cit. on pp. 33, 50).

[91] Y. Labyed and T. A. Bigelow. “Estimating the total ultrasound attenuation along the propagation path by using a reference phantom”. In: *The Journal of the Acoustical Society of America* 128.5 (Nov. 2010), pp. 3232–3238. DOI: 10.1121/1.3483739 (cit. on p. 62).

[92] I. Laina, N. Rieke, C. Rupprecht, J. P. Vizcaíno, A. Eslami, F. Tombari, and N. Navab. “Concurrent Segmentation and Localization for Tracking of Surgical Instruments”. In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2017*. Ed. by M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. L. Collins, and S. Duchesne. Vol. 10434. 00000. Cham: Springer International Publishing, 2017, pp. 664–672. DOI: 10.1007/978-3-319-66185-8_75 (cit. on pp. 34, 42).

[93] F. Lalys and P. Jannin. “Surgical process modelling: a review”. In: *International Journal of Computer Assisted Radiology and Surgery* 9.3 (May 2014). 00096, pp. 495–511. DOI: 10.1007/s11548-013-0940-5 (cit. on p. 39).

[94] R. J. Landreneau, M. J. Mack, S. R. Hazelrigg, R. D. Dowling, T. E. Acuff, M. J. Magee, and P. F. Ferson. “Video-assisted thoracic surgery: Basic technical concepts and intercostal approach strategies”. In: *The Annals of Thoracic Surgery* 54.4 (Oct. 1992). 00474, pp. 800–807. DOI: 10.1016/0003-4975(92)91040-G (cit. on pp. 7, 9).

[95] A. Lang, P. Mousavi, G. Fichtinger, and P. Abolmaesumi. “Fusion of electromagnetic tracking with speckle-tracked 3D freehand ultrasound using an unscented Kalman filter”. In: SPIE Medical Imaging. Ed. by S. A. McAleavey and J. D’hooge. Lake Buena Vista, FL, Feb. 26, 2009, 72651A. DOI: 10.1117/12.813879 (cit. on pp. 95, 96).

[96] J. E. Lang, S. Mannava, A. J. Floyd, M. S. Goddard, B. P. Smith, A. Mofidi, T. M. Seyler, and R. H. Jinnah. “Robotic systems in orthopaedic surgery”. In: *The Journal of Bone and Joint Surgery. British volume* 93-B.10 (Oct. 2011). 00087, pp. 1296–1299. DOI: 10.1302/0301-620X.93B10.27418 (cit. on pp. 47, 48).

[97] A. Lasso, T. Heffter, A. Rankin, C. Pinter, T. Ungi, and G. Fichtinger. “PLUS: Open-Source Toolkit for Ultrasound-Guided Intervention Systems”. In: *IEEE Transactions on Biomedical Engineering* 61.10 (Oct. 2014). 00150, pp. 2527–2537. DOI: 10.1109/TBME.2014.2322864 (cit. on p. 70).

[98] G. I. Lee, M. R. Lee, T. Clanton, E. Sutton, A. E. Park, and M. R. Marohn. “Comparative assessment of physical and cognitive ergonomics associated with robotic and traditional laparoscopic surgeries”. In: *Surgical Endoscopy* 28.2 (Feb. 2014). 00055, pp. 456–465. DOI: 10.1007/s00464-013-3213-z (cit. on p. 42).

[99] G. B. Levi Sandri, E. de Werra, G. Mascianà, F. Guerra, G. Spoletini, and Q. Lai. “The use of robotic surgery in abdominal organ transplantation: A literature review”. In: *Clinical Transplantation* 31.1 (Jan. 2017). 00009, e12856. DOI: 10.1111/ctr.12856 (cit. on p. 50).

[100] N. D. Lewis. “The scientific method in medicine”. In: *J Natl Med Assoc* 50.5 (Sept. 1958). 00001, pp. 325–328 (cit. on p. 6).

[101] G. Li, D. Citrin, K. Camphausen, B. Mueller, C. Burman, B. Mychalczak, R. W. Miller, and Y. Song. “Advances in 4D Medical Imaging and 4D Radiation Therapy”. In: *Technology in Cancer Research & Treatment* 7.1 (Feb. 2008). 00163, pp. 67–81. DOI: 10.1177/153303460800700109 (cit. on p. 18).

[102] K. H. C. Li, C. Kui, E. K. M. Lee, C. S. Ho, S. H. Sunny Hei, W. Wu, W. T. Wong, J. Voll, G. Li, T. Liu, B. Yan, J. Chan, G. Tse, and I. D. Keenan. “The role of 3D printing in anatomy education and surgical training: A narrative review”. In: *MedEdPublish* 6.2 (2017). DOI: 10.15694/mep.2017.000092 (cit. on p. 36).

[103] R. Londei, M. Esposito, B. Diotte, S.Weidert, E. Euler, P. Thaller, N. Navab, and P. Fallavollita. “Intra-operative augmented reality in distal locking”. In: *IJCARS 2015*. 2015 (cit. on p. 126).

[104] R. Londei, M. Esposito, B. Diotte, S.Weidert, E. Euler, P. Thaller, N. Navab, and P. Fallavollita. “The Augmented Circles: A Video-Guided Solution for the Down-the-Beam Positioning of IM Nail Holes”. In: *IPCAI 2014*. 2014 (cit. on p. 126).

[105] N. Loy Rodas and N. Padoy. “Seeing is believing: increasing intraoperative awareness to scattered radiation in interventional procedures by combining augmented reality, Monte Carlo simulations and wireless dosimeters”. In: *International Journal of Computer Assisted Radiology and Surgery* 10.8 (Aug. 2015), pp. 1181–1191. DOI: 10.1007/s11548-015-1161-x (cit. on p. 33).

[106] E. Lugez, H. Sadjadi, C. P. Joshi, S. G. Akl, and G. Fichtinger. “Improved electromagnetic tracking for catheter path reconstruction with application in high-dose-rate brachytherapy”. In: *International Journal of Computer Assisted Radiology and Surgery* 12.4 (Apr. 2017). 00003, pp. 681–689. DOI: 10.1007/s11548-017-1534-4 (cit. on p. 96).

[107] M. J. Mack. “Minimally Invasive and Robotic Surgery”. In: *JAMA* 285.5 (Feb. 7, 2001), p. 568. DOI: 10.1001/jama.285.5.568 (cit. on pp. 10, 33).

[108] L. Maier-Hein, P. Mountney, A. Bartoli, H. Elhawary, D. Elson, A. Groch, A. Kolb, M. Rodrigues, J. Sorger, S. Speidel, and D. Stoyanov. “Optical techniques for 3D surface reconstruction in computer-assisted laparoscopic surgery”. In: *Medical Image Analysis* 17.8 (Dec. 2013), pp. 974–996. DOI: 10.1016/j.media.2013.04.003 (cit. on p. 36).

[109] L. Maier-Hein, S. S. Vedula, S. Speidel, N. Navab, R. Kikinis, A. Park, M. Eisenmann, H. Feussner, G. Forestier, S. Giannarou, M. Hashizume, D. Katic, H. Kenngott, M. Kranzfelder, A. Malpani, K. März, T. Neumuth, N. Padoy, C. Pugh, N. Schoch, D. Stoyanov, R. Taylor, M. Wagner, G. D. Hager, and P. Jannin. “Surgical data science for next-generation interventions”. In: *Nature Biomedical Engineering* 1.9 (Sept. 2017). 00015, pp. 691–696. DOI: 10.1038/s41551-017-0132-7 (cit. on p. 40).

[110] J. Maintz and M. A. Viergever. “A survey of medical image registration”. In: *Medical Image Analysis* 2.1 (Mar. 1998). 04262, pp. 1–36. DOI: 10.1016/S1361-8415(01)80026-8 (cit. on p. 30).

[111] E. Marchand, F. Spindler, and F. Chaumette. “ViSP for visual servoing: a generic software platform with a wide class of robot control skills”. In: *IEEE Robotics & Automation Magazine* 12.4 (Dec. 2005). 00343, pp. 40–52. DOI: 10.1109/MRA.2005.1577023 (cit. on p. 76).

[112] S. Matinfar, M. A. Nasseri, U. Eck, H. Roodaki, N. Navab, C. P. Lohmann, M. Maier, and N. Navab. “Surgical Soundtracks: Towards Automatic Musical Augmentation of Surgical Procedures”. In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2017*. Ed. by M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. L. Collins, and S. Duchesne. Vol. 10434. Cham: Springer International Publishing, 2017, pp. 673–681. DOI: 10.1007/978-3-319-66185-8_76 (cit. on p. 37).

[113] T. A. Mattei, A. H. Rodriguez, D. Sambhara, and E. Mendel. “Current state-of-the-art and future perspectives of robotic technology in neurosurgery”. In: *Neurosurgical Review* 37.3 (July 2014). 00032, pp. 357–366. DOI: 10.1007/s10143-014-0540-z (cit. on pp. 45, 47, 49, 54).

[114] P. Matthies, J. Gardiazabal, A. Okur, J. Vogel, T. Lasser, and N. Navab. “Mini gamma cameras for intra-operative nuclear tomographic reconstruction”. In: *Medical Image Analysis* 18.8 (Dec. 2014), pp. 1329–1336. DOI: 10.1016/j.media.2014.04.009 (cit. on p. 78).

[115] P. Matthies, K. Sharma, A. Okur, J. Gardiazabal, J. Vogel, T. Lasser, and N. Navab. “First Use of Mini Gamma Cameras for Intra-operative Robotic SPECT Reconstruction”. In: *Advanced Information Systems Engineering*. Ed. by C. Salinesi, M. C. Norrie, and Ó. Pastor. Red. by D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, and G. Weikum. Vol. 7908. 00000. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 163–170. DOI: 10.1007/978-3-642-40811-3_21 (cit. on p. 28).

[116] I. McClure. “Bad Medicine: Doctors Doing Harm Since Hippocrates”. In: *BMJ* 333.7568 (Sept. 16, 2006). 00001, p. 606. DOI: 10.1136/bmj.333.7568.606 (cit. on p. 120).

[117] H. M. Mentis, A. Chellali, K. Manser, C. G. L. Cao, and S. D. Schwaitzberg. “A systematic review of the effect of distraction on surgeon performance: directions for operating room policy and surgical training”. In: *Surgical Endoscopy* 30.5 (May 2016). 00020, pp. 1713–1724. DOI: 10.1007/s00464-015-4443-z (cit. on p. 42).

[118] R. Mercer, C. P. Marcella, D. K. Carney, and R. W. McDonald. “Occupational health hazards to the ultrasonographer and their possible prevention”. In: *Journal of the American Society of Echocardiography* 10.4 (May 1997), pp. 363–366. DOI: 10.1016/S0894-7317(97)70074-9 (cit. on p. 26).

[119] D. Miller, C. Lippert, F. Vollmer, O. Bozinov, L. Benes, D. Schulte, and U. Sure. “Comparison of different reconstruction algorithms for three-dimensional ultrasound imaging in a neurosurgical setting: Comparison 3D US reconstruction algorithms”. In: *The International Journal of Medical Robotics and Computer Assisted Surgery* 8.3 (Sept. 2012). 00000, pp. 348–359. DOI: 10.1002/rcs.1420 (cit. on p. 98).

[120] M. D. Miller, B. J. Cole, and Miller/Cole, eds. *Textbook of arthroscopy*. OCLC: 636609980. Philadelphia, Pa: Saunders, 2004. 862 pp. (cit. on pp. 7, 9).

[121] F. Milletari, N. Rieke, M. Baust, M. Esposito, and N. Navab. “CFCM: Segmentation via Coarse to Fine Context Memory”. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. Ed. by A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger. Vol. 11073. 00000. Cham: Springer International Publishing, 2018, pp. 667–674. DOI: 10.1007/978-3-030-00937-3_76 (cit. on p. 125).

[122] D. L. Mills. *Computer network time synchronization: the Network Time Protocol*. 00221 OCLC: ocm62282681. Boca Raton, FL: CRC/Taylor & Francis, 2006. 286 pp. (cit. on p. 75).

[123] M. Moakher. “Means and Averaging in the Group of Rotations”. In: *SIAM Journal on Matrix Analysis and Applications* 24.1 (Jan. 2002). 00383, pp. 1–16. DOI: 10.1137/S0895479801383877 (cit. on p. 140).

[124] O. Mohareri, J. Ischia, P. C. Black, C. Schneider, J. Lobo, L. Goldenberg, and S. E. Salcudean. “Intraoperative Registered Transrectal Ultrasound Guidance for Robot-Assisted Laparoscopic Radical Prostatectomy”. In: *The Journal of Urology* 193.1 (Jan. 2015). 00011, pp. 302–312. DOI: 10.1016/j.juro.2014.05.124 (cit. on p. 53).

[125] H. Moradi, S. Tang, and S. E. Salcudean. “Towards Intra-operative Prostate Photoacoustic Imaging: Configuration Evaluation and Implementation Using the da Vinci Research Kit”. In: *IEEE Transactions on Medical Imaging* (2018), pp. 1–1. DOI: 10.1109/TMI.2018.2855166 (cit. on p. 41).

[126] L. Morelli, S. Guadagni, G. Di Franco, M. Palmeri, G. Di Candio, and F. Mosca. “Da Vinci single site© surgical platform in clinical practice: a systematic review: Da Vinci single site© surgical platform in clinical practice”. In: *The International Journal of Medical Robotics and Computer Assisted Surgery* 12.4 (Dec. 2016), pp. 724–734. DOI: 10.1002/rcs.1713 (cit. on p. 47).

[127] G. E. Morgan, M. S. Mikhail, and M. J. Murray. *Clinical anesthesiology*. 4th ed. 01589 OCLC: ocm62286575. New York: Lange Medical Books/McGraw Hill, Medical Pub. Division, 2006. 1105 pp. (cit. on p. 6).

[128] G. P. Moustris, S. C. Hiridis, K. M. Deliparaschos, and K. M. Konstantinidis. “Evolution of autonomous and semi-autonomous robotic surgical systems: a review of the literature”. In: *The International Journal of Medical Robotics and Computer Assisted Surgery* 7.4 (Dec. 2011). 00199, pp. 375–392. DOI: 10.1002/rcs.408 (cit. on p. 121).

[129] M. H. Mozaffari and W.-S. Lee. “Freehand 3-D Ultrasound Imaging: A Systematic Review”. In: *Ultrasound Med Biol* 43.10 (2017). 00004, pp. 2099–2124. DOI: 10.1016/j.ultrasmedbio.2017.06.009 (cit. on pp. 93, 98).

[130] S. Mukherjee. “A.I. Versus M.D.” In: *Annals of Medicine, The New Yorker* (April 3, 2017 Mar. 27, 2017). 00013 (cit. on p. 120).

[131] C. Nafis, V. Jensen, L. Beauregard, and P. Anderson. “Method for estimating dynamic EM tracking accuracy of surgical navigation tools”. In: Medical Imaging. Ed. by K. R. Cleary and R. L. Galloway Jr. 00088. San Diego, CA, Mar. 2, 2006, 61410K. DOI: 10.1117/12.653448 (cit. on p. 38).

[132] K. Nagpal, A. Vats, B. Lamb, H. Ashrafian, N. Sevdalis, C. Vincent, and K. Moorthy. “Information Transfer and Communication in Surgery: A Systematic Review”. In: *Annals of Surgery* 252.2 (Aug. 2010). 00164, pp. 225–239. DOI: 10.1097/SLA.0b013e3181e495c2 (cit. on p. 39).

[133] Y. Nakamura, H. Hanafusa, and T. Yoshikawa. “Task-Priority Based Redundancy Control of Robot Manipulators”. In: *The International Journal of Robotics Research* 6.2 (June 1987). 00912, pp. 3–15. DOI: 10.1177/027836498700600201 (cit. on p. 68).

[134] M. A. Nasseri, M. Eder, S. Nair, E. C. Dean, M. Maier, D. Zapp, C. P. Lohmann, and A. Knoll. “The introduction of a new robot for assistance in ophthalmic surgery”. In: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 00032. Osaka: IEEE, July 2013, pp. 5682–5685. DOI: 10.1109/EMBC.2013.6610840 (cit. on p. 54).

[135] N. Nathoo, M. C. Çavuşoğlu, M. A. Vogelbaum, and G. H. Barnett. “In Touch with Robotics: Neurosurgery for the Future”. In: *Neurosurgery* 56.3 (Mar. 2005). 00153, pp. 421–433. DOI: 10.1227/01.NEU.0000153929.68024.CF (cit. on pp. 45, 49–51).

[136] N. Navab, T. Blum, L. Wang, A. Okur, and T. Wendler. “First Deployments of Augmented Reality in Operating Rooms”. In: *Computer* 45.7 (July 2012). 00072, pp. 48–55. DOI: 10.1109/MC.2012.75 (cit. on pp. 19, 34, 35).

[137] N. Navab, M. Fellow, C. Hennersperger, B. Frisch, and B. Fürst. “Personalized, relevance-based Multimodal Robotic Imaging and augmented reality for Computer Assisted Interventions”. In: *Medical Image Analysis* 33 (Oct. 2016). 00013, pp. 64–71. DOI: 10.1016/j.media.2016.06.021 (cit. on p. 40).

[138] B. J. Nelson, I. K. Kaliakatsos, and J. J. Abbott. “Microrobots for Minimally Invasive Medicine”. In: *Annual Review of Biomedical Engineering* 12.1 (July 2010). 00876, pp. 55–85. DOI: 10.1146/annurev-bioeng-010510-103409 (cit. on p. 54).

[139] J. Noble and D. Boukerroui. “Ultrasound image segmentation: a survey”. In: *IEEE Transactions on Medical Imaging* 25.8 (Aug. 2006). 01005, pp. 987–1010. DOI: 10.1109/TMI.2006.877092 (cit. on p. 91).

[140] R. Obwegeser, K. Lorenz, M. Hohlagschwandtner, K. Czerwenka, B. Schneider, and E. Kubista. “Axillary lymph nodes in breast cancer: is size related to metastatic involvement?” In: *World J Surg* 24.5 (May 2000). 00046, pp. 546–550 (cit. on pp. 59, 82).

[141] A. M. Okamura. “Haptic feedback in robot-assisted minimally invasive surgery:” in: *Current Opinion in Urology* 19.1 (Jan. 2009). 00333, pp. 102–107. DOI: 10.1097/MOU.0b013e32831a478c (cit. on p. 54).

[142] A. Okur, C. Hennersperger, B. Runyan, J. Gardiazabal, M. Keicher, S. Paepke, T. Wendler, and N. Navab. “FhSPECT-US guided needle biopsy of sentinel lymph nodes in the axilla: is it feasible?” In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2014, pp. 577–584 (cit. on pp. 28, 61).

[143] F. P. Oliveira and J. M. R. Tavares. “Medical image registration: a review”. In: *Computer Methods in Biomechanics and Biomedical Engineering* 17.2 (Jan. 25, 2014). 00359, pp. 73–93. DOI: 10.1080/10255842.2012.670855 (cit. on p. 36).

[144] S. Ourselin, M. Emberton, and T. Vercauteren. “From computer-assisted intervention research to clinical impact: The need for a holistic approach”. In: *Med Image Anal* 33 (2016). 00003, pp. 72–78. DOI: 10.1016/j.media.2016.06.018 (cit. on pp. 3, 41).

[145] M. Pamilo, M. Soiva, and E. M. Lavast. “Real-time ultrasound, axillary mammography, and clinical examination in the detection of axillary lymph node metastases in breast cancer patients.” In: *Journal of Ultrasound in Medicine* 8.3 (Mar. 1989). 00187, pp. 115–120. DOI: 10.7863/jum.1989.8.3.115 (cit. on p. 59).

[146] R. Pani, R. Pellegrini, M. Cinti, P. Bennati, A. Fabbri, S. Ridolfi, R. Scafè, G. De Vincentis, E. Di Castro, N. Polli, M. Caratozzolo, M. Mattioli, P. Boccaccio, G. Moschini, N. Lanconelli, S. Lo Meo, F. Navarria, D. Sacco, V. Cencelli, T. Baroncelli, and F. de Notaristefani. "Dual Modality Ultrasound-SPET Detector for Molecular Imaging". In: *Nuclear Physics B - Proceedings Supplements* 215.1 (June 2011), pp. 319–323. DOI: 10.1016/j.nuclphysbps.2011.04.043 (cit. on p. 61).

[147] A. Park, G. Lee, F. J. Seagull, N. Meenaghan, and D. Dexter. "Patients Benefit While Surgeons Suffer: An Impending Epidemic". In: *Journal of the American College of Surgeons* 210.3 (Mar. 2010). 00248, pp. 306–313. DOI: 10.1016/j.jamcollsurg.2009.10.017 (cit. on pp. 33, 42).

[148] F. C. Park and R. W. Brockett. "Kinematic Dexterity of Robotic Mechanisms". In: *The International Journal of Robotics Research* 13.1 (Feb. 1994). 00210, pp. 1–15. DOI: 10.1177/027836499401300101 (cit. on p. 101).

[149] K. R. Parks and E. J. Hagopian. "Introduction: The Importance of Ultrasound in a Surgical Practice". In: *Abdominal Ultrasound for Surgeons*. Ed. by E. J. Hagopian and J. Machi. 00004. New York, NY: Springer New York, 2014, pp. 3–6. DOI: 10.1007/978-1-4614-9599-4_1 (cit. on p. 90).

[150] N. Patronik, C. Riviere, S. El Qarra, and M. Zenati. "The HeartLander: A novel epicardial crawling robot for myocardial injections". In: *International Congress Series* 1281 (May 2005). 00019, pp. 735–739. DOI: 10.1016/j.ics.2005.03.325 (cit. on p. 54).

[151] T. M. Peters and C. A. Linte. "Image-guided interventions and computer-integrated therapy: Quo vadis?" In: *Medical Image Analysis* 33 (Oct. 2016), pp. 56–63. DOI: 10.1016/j.media.2016.06.004 (cit. on p. 41).

[152] T. Pock and A. Chambolle. "Diagonal preconditioning for first order primal-dual algorithms in convex optimization". In: *2011 International Conference on Computer Vision*. 2011 IEEE International Conference on Computer Vision (ICCV). 00301. Barcelona, Spain: IEEE, Nov. 2011, pp. 1762–1769. DOI: 10.1109/ICCV.2011.6126441 (cit. on p. 102).

[153] C. Polito, R. Pellegrini, M. Cinti, G. De Vincentis, S. Lo Meo, A. Fabbri, P. Bennati, V. O. Cencelli, and R. Pani. "Dual-modality imaging with a ultrasound-gamma device for oncology". In: *Radiation Physics and Chemistry* 147 (June 2018). 00001, pp. 77–84. DOI: 10.1016/j.radphyschem.2018.02.018 (cit. on p. 61).

[154] R. W. Prager, U. Z. Ijaz, A. H. Gee, and G. M. Treece. "Three-dimensional ultrasound imaging". In: *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* 224.2 (Feb. 2010), pp. 193–223. DOI: 10.1243/09544119JEIM586 (cit. on p. 91).

[155] R. Prevost, M. Salehi, J. Sprung, A. Ladikos, R. Bauer, and W. Wein. "Deep Learning for Sensorless 3D Freehand Ultrasound Imaging". In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2017*. Ed. by M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. L. Collins, and S. Duchesne. Vol. 10434. Cham: Springer International Publishing, 2017, pp. 628–636. DOI: 10.1007/978-3-319-66185-8_71 (cit. on pp. 89, 93).

[156] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng. "ROS: an open-source Robot Operating System". In: *ICRA workshop on open source software*. Vol. 3. 04959. Kobe, Japan, 2009, p. 5 (cit. on p. 71).

[157] D. Rattner and A. Kalloo. "ASGE/SAGES Working Group on Natural Orifice Translumenal Endoscopic Surgery: October 2005". In: *Surgical Endoscopy* 20.2 (Feb. 2006), pp. 329–333. DOI: 10.1007/s00464-005-3006-0 (cit. on p. 50).

[158] F. Rodriguez y Baena and B. Davies. "Robotic surgery: from autonomous systems to intelligent tools". In: *Robotica* 28.2 (Mar. 2010). 00027, p. 163. DOI: 10.1017/S0263574709990427 (cit. on p. 121).

[159] F. J. Romero-Ramirez, R. Muñoz-Salinas, and R. Medina-Carnicer. "Speeded up detection of squared fiducial markers". In: *Image and Vision Computing* 76 (Aug. 2018), pp. 38–47. DOI: 10.1016/j.imavis.2018.05.004 (cit. on p. 63).

[160] A. Rosenfeld, R. Zemel, and J. K. Tsotsos. “The Elephant in the Room”. In: *arXiv:1808.03305 [cs]* (Aug. 9, 2018). 00002. arXiv: 1808.03305 (cit. on p. 39).

[161] H. Sadjadi, K. Hashtrudi-Zaad, and G. Fichtinger. “Simultaneous localization and calibration for electromagnetic tracking systems: Simultaneous Localization and EM Calibration”. In: *The International Journal of Medical Robotics and Computer Assisted Surgery* 12.2 (June 2016). 00000, pp. 189–198. DOI: 10.1002/rcs.1670 (cit. on p. 96).

[162] R. San José-Estépar, M. Martín-Fernández, P. Caballero-Martínez, C. Alberola-López, and J. Ruiz-Alzola. “A theoretical framework to three-dimensional ultrasound reconstruction from irregularly sampled data”. In: *Ultrasound in Medicine & Biology* 29.2 (Feb. 2003). 00000, pp. 255–269. DOI: 10.1016/S0301-5629(02)00762-7 (cit. on p. 98).

[163] A. Savitzky and M. J. E. Golay. “Smoothing and Differentiation of Data by Simplified Least Squares Procedures.” In: *Analytical Chemistry* 36.8 (July 1964). 13342, pp. 1627–1639. DOI: 10.1021/ac60214a047 (cit. on p. 109).

[164] C. Schmalz, F. Forster, A. Schick, and E. Angelopoulou. “An endoscopic 3D scanner based on structured light”. In: *Medical Image Analysis* 16.5 (July 2012), pp. 1063–1072. DOI: 10.1016/j.media.2012.04.001 (cit. on p. 35).

[165] A. Schmidt. “Implicit human computer interaction through context”. In: *Personal Technologies* 4.2 (June 2000), pp. 191–199. DOI: 10.1007/BF01324126 (cit. on p. 52).

[166] B. Schuller, S. Can, H. Feussner, M. Wollmer, D. Arisc, and B. Hornler. “Speech control in surgery: A field analysis and strategies”. In: *2009 IEEE International Conference on Multimedia and Expo*. 2009 IEEE International Conference on Multimedia and Expo (ICME). 00009. New York, NY, USA: IEEE, June 2009, pp. 1214–1217. DOI: 10.1109/ICME.2009.5202719 (cit. on pp. 38, 52).

[167] B. Schuller, G. Rigoll, S. Can, and H. Feussner. “Emotion sensitive speech control for human-robot interaction in minimal invasive surgery”. In: *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*. 2008 RO-MAN: The 17th IEEE International Symposium on Robot and Human Interactive Communication. 00028. Munich, Germany: IEEE, Aug. 2008, pp. 453–458. DOI: 10.1109/ROMAN.2008.4600708 (cit. on p. 52).

[168] C. Schulte zu Berge, M. Baust, A. Kapoor, and N. Navab. “Predicate-Based Focus-and-Context Visualization for 3D Ultrasound”. In: *IEEE Transactions on Visualization and Computer Graphics* 20.12 (Dec. 31, 2014), pp. 2379–2387. DOI: 10.1109/TVCG.2014.2346317 (cit. on p. 36).

[169] C. Schulte zu Berge, A. Grunau, H. Mahmud, and N. Navab. {*CAMPVis*} – {*A*} *Game Engine-inspired Research Framework for Medical Imaging and Visualization*. Technische Universität München, 2014 (cit. on pp. 70, 78).

[170] S. Seshamani, W. Lau, and G. Hager. “Real-Time Endoscopic Mosaicking”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006*. Ed. by R. Larsen, M. Nielsen, and J. Sporring. Red. by D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, and G. Weikum. Vol. 4190. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 355–363. DOI: 10.1007/11866565_44 (cit. on p. 36).

[171] M. Shoham, M. Burman, E. Zehavi, L. Joskowicz, E. Batkilin, and Y. Kunicher. “Bone-mounted miniature robot for surgical procedures: concept and clinical applications”. In: *IEEE Transactions on Robotics and Automation* 19.5 (Oct. 2003). 00310, pp. 893–901. DOI: 10.1109/TRA.2003.817075 (cit. on p. 54).

[172] R. L. Siegel, K. D. Miller, and A. Jemal. “Cancer statistics, 2017”. In: *CA: A Cancer Journal for Clinicians* 67.1 (Jan. 2017). 00674, pp. 7–30. DOI: 10.3322/caac.21387 (cit. on p. 58).

[173] T. Sielhorst, M. Feuerstein, and N. Navab. “Advanced Medical Displays: A Literature Review of Augmented Reality”. In: *Journal of Display Technology* 4.4 (Dec. 2008). 00193, pp. 451–467. DOI: 10.1109/JDT.2008.2001575 (cit. on pp. 36, 42).

[174] G. Singer. “Occupational radiation exposure to the surgeon”. In: *J Am Acad Orthop Surg* 13.1 (Feb. 2005). 00185, pp. 69–76 (cit. on p. 23).

[175] O. V. Solberg, F. Lindseth, H. Torp, R. E. Blake, and T. A. Nagelhus Hernes. “Freehand 3D Ultrasound Reconstruction Algorithms—A Review”. In: *Ultrasound in Medicine & Biology* 33.7 (July 2007). 00198, pp. 991–1009. DOI: 10.1016/j.ultrasmedbio.2007.02.015 (cit. on p. 98).

[176] C. Song, P. L. Gehlbach, and J. U. Kang. “Active tremor cancellation by a “Smart” handheld vitreoretinal microsurgical tool using swept source optical coherence tomography”. In: *Optics Express* 20.21 (Oct. 8, 2012). 00053, p. 23414. DOI: 10.1364/OE.20.023414 (cit. on p. 49).

[177] A. Sotiras, C. Davatzikos, and N. Paragios. “Deformable Medical Image Registration: A Survey”. In: *IEEE Transactions on Medical Imaging* 32.7 (July 2013). 00854, pp. 1153–1190. DOI: 10.1109/TMI.2013.2265603 (cit. on p. 36).

[178] S. J. Spaner and G. L. Warnock. “A Brief History of Endoscopy, Laparoscopy, and Laparoscopic Surgery”. In: *Journal of Laparoendoscopic & Advanced Surgical Techniques* 7.6 (Dec. 1997). 00188, pp. 369–373. DOI: 10.1089/lap.1997.7.369 (cit. on p. 7).

[179] A. Taruttis and V. Ntziachristos. “Advances in real-time multispectral optoacoustic imaging and its applications”. In: *Nature Photonics* 9.4 (2015). 00197, p. 219 (cit. on p. 29).

[180] J. J. Tate, V. Lewis, T. Archer, P. G. Guyer, G. T. Royle, and I. Taylor. “Ultrasound detection of axillary lymph node metastases in breast cancer”. In: *Eur J Surg Oncol* 15.2 (Apr. 1989). 00107, pp. 139–141 (cit. on p. 59).

[181] R. Taylor. “A Perspective on Medical Robotics”. In: *Proceedings of the IEEE* 94.9 (Sept. 2006). 00181, pp. 1652–1664. DOI: 10.1109/JPROC.2006.880669 (cit. on pp. 45, 47).

[182] R. Taylor and D. Stoianovici. “Medical robotics in computer-integrated surgery”. In: *IEEE Transactions on Robotics and Automation* 19.5 (Oct. 2003). 00000, pp. 765–781. DOI: 10.1109/TRA.2003.817058 (cit. on pp. 38, 45, 47, 54).

[183] R. Taylor, P. Jensen, L. Whitcomb, A. Barnes, R. Kumar, D. Stoianovici, P. Gupta, Z. Wang, E. Dejuan, and L. Kavoussi. “A Steady-Hand Robotic System for Microsurgical Augmentation”. In: *The International Journal of Robotics Research* 18.12 (Dec. 1999). 00557, pp. 1201–1210. DOI: 10.1177/02783649922067807 (cit. on p. 49).

[184] R. H. Taylor and P. Kazanzides. “Medical Robotics and Computer-Integrated Interventional Medicine”. In: *Biomedical Information Technology*. Elsevier, 2008, pp. 393–416. DOI: 10.1016/B978-012373583-6.50022-0 (cit. on p. 38).

[185] “The essential physics of medical imaging. Second edition. 2002”. In: *European Journal of Nuclear Medicine and Molecular Imaging* 30.12 (Dec. 1, 2003). 00000, pp. 1713–1713. DOI: 10.1007/s00259-003-1310-2 (cit. on pp. 22, 23).

[186] J. Tokuda, G. S. Fischer, X. Papademetris, Z. Yaniv, L. Ibanez, P. Cheng, H. Liu, J. Blevins, J. Arata, A. J. Golby, T. Kapur, S. Pieper, E. C. Burdette, G. Fichtinger, C. M. Tempany, and N. Hata. “OpenIGTLink: an open network protocol for image-guided therapy environment”. In: *Int J Med Robot* 5.4 (Dec. 2009). 00205, pp. 423–434. DOI: 10.1002/rcs.274 (cit. on p. 70).

[187] M. Tonutti, D. S. Elson, G.-Z. Yang, A. W. Darzi, and M. H. Sodergren. “The role of technology in minimally invasive surgery: state of the art, recent developments and future directions”. In: *Postgraduate Medical Journal* 93.1097 (Mar. 2017). 00009, pp. 159–167. DOI: 10.1136/postgradmedj-2016-134311 (cit. on pp. 7, 33).

[188] G. M. Treece, A. H. Gee, R. W. Prager, C. J. Cash, and L. H. Berman. “High-definition freehand 3-D ultrasound”. In: *Ultrasound in Medicine & Biology* 29.4 (Apr. 2003). 00148, pp. 529–546. DOI: 10.1016/S0301-5629(02)00735-4 (cit. on pp. 26, 92, 93).

[189] J. W. Trobaugh, D. J. Trobaugh, and W. D. Richard. “Three-dimensional imaging with stereotactic ultrasonography”. In: *Computerized Medical Imaging and Graphics* 18.5 (Sept. 1994). 00102, pp. 315–323. DOI: 10.1016/0895-6111(94)90002-7 (cit. on p. 93).

[190] R. Tsai and R. Lenz. “A new technique for fully autonomous and efficient 3D robotics hand/eye calibration”. In: *IEEE Transactions on Robotics and Automation* 5.3 (June 1989). 01009, pp. 345–358. DOI: 10.1109/70.34770 (cit. on p. 76).

[191] H. Tutkun Sen and P. Kazanzides. “Bayesian filtering to improve the dynamic accuracy of electromagnetic tracking”. In: *2013 IEEE International Symposium on Robotic and Sensors Environments (ROSE)*. 2013 IEEE International Symposium on Robotic and Sensors Environments (ROSE). 00003. Washington, DC, USA: IEEE, Oct. 2013, pp. 90–95. DOI: 10.1109/ROSE.2013.6698424 (cit. on p. 97).

[192] A. Vaccarella, E. De Momi, A. Enquobahrie, and G. Ferrigno. “Unscented Kalman Filter Based Sensor Fusion for Robust Optical and Electromagnetic Tracking in Surgical Navigation”. In: *IEEE Transactions on Instrumentation and Measurement* 62.7 (July 2013). 00043, pp. 2067–2081. DOI: 10.1109/TIM.2013.2248304 (cit. on p. 96).

[193] P. Valagussa, G. Bonadonna, and U. Veronesi. “Patterns of relapse and survival following radical mastectomy.Analysis of 716 consecutive patients”. In: *Cancer* 41.3 (Mar. 1978). 00614, pp. 1170–1178. DOI: 10.1002/1097-0142(197803)41:3<1170::AID-CNCR2820410355>3.0.CO;2-I (cit. on p. 58).

[194] S. Vidal-Sicart and R. Valdés Olmos. “Sentinel Node Mapping for Breast Cancer: Current Situation”. In: *Journal of Oncology* 2012 (2012). 00044, pp. 1–7. DOI: 10.1155/2012/361341 (cit. on p. 59).

[195] S. Virga, O. Zettinig, M. Esposito, K. Pfister, B. Frisch, T. Neff, N. Navab, and C. Hennersperger. “Automatic force-compliant robotic ultrasound screening of abdominal aortic aneurysms”. In: *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*. IEEE, 2016, pp. 508–513 (cit. on pp. 49, 125).

[196] V. Vitiello, Su-Lin Lee, T. P. Cundy, and Guang-Zhong Yang. “Emerging Robotic Platforms for Minimally Invasive Surgery”. In: *IEEE Reviews in Biomedical Engineering* 6 (2013), pp. 111–126. DOI: 10.1109/RBME.2012.2236311 (cit. on p. 45).

[197] S. Voros, G.-P. Haber, J.-F. Menudet, J.-A. Long, and P. Cinquin. “ViKY Robotic Scope Holder: Initial Clinical Experience and Preliminary Results Using Instrument Tracking”. In: *IEEE/ASME Transactions on Mechatronics* (Dec. 2010). 00044. DOI: 10.1109/TMECH.2010.2080683 (cit. on p. 38).

[198] R. Walsh, M. Soehl, A. Rankin, A. Lasso, and G. Fichtinger. “Design of a tracked ultrasound calibration phantom made of LEGO bricks”. In: SPIE Medical Imaging. Ed. by Z. R. Yaniv and D. R. Holmes. 00000. San Diego, California, USA, Mar. 12, 2014, p. 90362C. DOI: 10.1117/12.2043533 (cit. on p. 105).

[199] A. Weinmann, L. Demaret, and M. Storath. “Total Variation Regularization for Manifold-Valued Data”. In: *SIAM Journal on Imaging Sciences* 7.4 (Jan. 2014). 00050, pp. 2226–2257. DOI: 10.1137/130951075 (cit. on pp. 101, 103, 140).

[200] J. Welch, J. Johnson, M. Bax, R. Badr, and R. Shahidi. “A real-time freehand 3D ultrasound system for image-guided surgery”. In: *2000 IEEE Ultrasonics Symposium. Proceedings. An International Symposium (Cat. No.00CH37121)*. 2000 IEEE Ultrasonics Symposium. Proceedings. An International Symposium. Vol. 2. 00040. San Juan, Puerto Rico: IEEE, 2000, pp. 1601–1604. DOI: 10.1109/ULTSYM.2000.921630 (cit. on p. 92).

[201] E. P. Westebring – van der Putten, R. H. M. Goossens, J. J. Jakimowicz, and J. Dankelman. “Haptics in minimally invasive surgery – a review”. In: *Minimally Invasive Therapy & Allied Technologies* 17.1 (Jan. 2008), pp. 3–16. DOI: 10.1080/13645700701820242 (cit. on p. 54).

[202] D. A. Wiegmann, A. W. ElBardissi, J. A. Dearani, R. C. Daly, and T. M. Sundt. “Disruptions in surgical flow and their relationship to surgical errors: An exploratory investigation”. In: *Surgery* 142.5 (Nov. 2007). 00336, pp. 658–665. DOI: 10.1016/j.surg.2007.07.034 (cit. on p. 42).

[203] S. Wojcinski, Farrokh, Hille, Wiskirchen, Gyapong, Soliman, Degenhardt, and Hillemanns. “The Automated Breast Volume Scanner (ABVS): initial experiences in lesion detection compared with conventional handheld B-mode ultrasound: a pilot study of 50 cases”. In: *International Journal of Women’s Health* (Oct. 2011). 00107, p. 337. DOI: 10.2147/IJWH.S23918 (cit. on p. 25).

[204] J. P. Womack, D. T. Jones, and D. Roos. *The machine that changed the world: the story of lean production ; Toyota’s secret weapon in the global car wars that is revolutionizing world industry*. 1. pb. ed. Business. 00057 OCLC: 85814817. New York, NY: Free Press, 2007. 339 pp. (cit. on p. 121).

[205] G.-Z. Yang, J. Cambias, K. Cleary, E. Daimler, J. Drake, P. E. Dupont, N. Hata, P. Kazanzides, S. Martel, R. V. Patel, V. J. Santos, and R. H. Taylor. “Medical robotics—Regulatory, ethical, and legal considerations for increasing levels of autonomy”. In: *Science Robotics* 2.4 (Mar. 15, 2017). 00015, eaam8638. DOI: 10.1126/scirobotics.aam8638 (cit. on p. 48).

[206] P. G. Yock and D. T. Linker. “Intravascular ultrasound. Looking below the surface of vascular disease.” In: *Circulation* 81.5 (May 1990). 00128, pp. 1715–1718. DOI: 10.1161/01.CIR.81.5.1715 (cit. on p. 26).

[207] S. Yoshinaga, K. Mabuchi, A. J. Sigurdson, M. M. Doody, and E. Ron. “Cancer Risks among Radiologists and Radiologic Technologists: Review of Epidemiologic Studies”. In: *Radiology* 233.2 (Nov. 2004). 00232, pp. 313–321. DOI: 10.1148/radiol.2332031119 (cit. on p. 23).

[208] S.-B. Yu and A. D. Watson. “Metal-Based X-ray Contrast Media”. In: *Chemical Reviews* 99.9 (Sept. 1999). 00347, pp. 2353–2378. DOI: 10.1021/cr980441p (cit. on p. 23).

[209] Yu Qi, H. Sadjadi, C. T. Yeo, K. Hashtrudi-Zaad, and G. Fichtinger. “Electromagnetic tracking performance analysis and optimization”. In: *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Chicago, IL: IEEE, Aug. 2014, pp. 6534–6538. DOI: 10.1109/EMBC.2014.6945125 (cit. on pp. 95, 96).

[210] E. Zacur, M. Bossa, and S. Olmos. “Left-Invariant Riemannian Geodesics on Spatial Transformation Groups”. In: *SIAM Journal on Imaging Sciences* 7.3 (Jan. 2014). 00008, pp. 1503–1557. DOI: 10.1137/130928352 (cit. on p. 101).

[211] M. Zefran, K. Vijay, and C. Croke. “Choice of Riemannian metrics for rigid body kinematics”. In: vol. 2. 00046. 1996 (cit. on p. 101).

[212] O. Zettinig, B. Frisch, S. Virga, M. Esposito, A. Rienmüller, B. Meyer, C. Hennersperger, Y.-M. Ryang, and N. Navab. “3D Ultrasound Registration-based Visual Servoing for Neurosurgical Navigation”. In: *International Journal of Computer Assisted Radiology and Surgery* in press (2017). DOI: 10.1007/s11548-017-1536-2 (cit. on p. 125).

[213] O. Zettinig, B. Fuerst, R. Kojcev, M. Esposito, M. Salehi, W. Wein, J. Rackerseder, E. Sinibaldi, B. Frisch, and N. Navab. “Toward real-time 3D ultrasound registration-based visual servoing for interventional navigation”. In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 945–950 (cit. on p. 126).

[214] S. Zhang. “Why An Autonomous Robot Won’t Replace Your Surgeon Anytime Soon”. In: *Wired* (May 4, 2016). 00002 (cit. on p. 120).

[215] Z. Zhang. “A flexible new technique for camera calibration”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.11 (Nov. 2000), pp. 1330–1334. DOI: 10.1109/34.888718 (cit. on pp. 64, 76).

Online Resources

[216] *'Words Fail Me. It's Insanity.' While Elon Musk Loves Tesla's New Tent Factory, Others Aren't So Sure*. Fortune. 00000. URL: http://fortune.com/2018/06/25/elon-musk-tesla-tent-factory-model-3/ (visited on Nov. 12, 2018) (cit. on p. 121).

[217] A. J. Hawkins. *Tesla relied on too many robots to build the Model 3, Elon Musk says*. The Verge. 00000. Apr. 13, 2018. URL: https://www.theverge.com/2018/4/13/17234296/tesla-model-3-robots-production-hell-elon-musk (visited on Nov. 12, 2018) (cit. on p. 121).

[218] *Opinion: No, AI Will Not Replace Radiologists*. The Scientist Magazine®. 00000. URL: https://www.the-scientist.com/news-opinion/opinion--no--ai-will-not-replace-radiologists-64506 (visited on Nov. 13, 2018) (cit. on p. 120).

[219] H. Stark. *Prepare Yourselves, Robots Will Soon Replace Doctors In Healthcare*. Forbes. 00000. URL: https://www.forbes.com/sites/haroldstark/2017/07/10/prepare-yourselves-robots-will-soon-replace-doctors-in-healthcare/ (visited on Nov. 12, 2018) (cit. on p. 120).

List of Figures

List of Tables