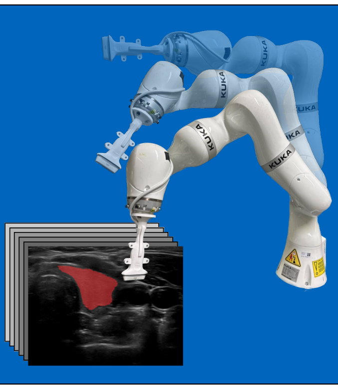
Computer Aided Medical Procedures
Prof. Dr. Nassir Navab



Dissertation

Intelligent, Applicable and Acceptable Robotic Ultrasound

Christine Eilers

School of Computation, Information and Technology
Technische Universität München

Technische Universität München
TUM School of Computation, Information and Technology

Intelligent, Applicable and Acceptable Robotic Ultrasound

Christine Eilers

Vollständiger Abdruck der von der TUM School of Computation, Information and Technology der Technischen Universität München zur Erlangung des akademischen Grades einer

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Alois Knoll

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

2. Prof. Dr. Pietro Valdastri

Die Dissertation wurde am 21.02.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Computation, Information and Technology am 13.07.2023 angenommen.

Christine Eilers
Intelligent, Applicable and Acceptable Robotic Ultrasound
Dissertation, Version 1.0

Technische Universität München
TUM School of Computation, Information and Technology
Lehrstuhl für Informatikanwendungen in der Medizin
Boltzmannstraße 3
85748 and Garching bei München

Abstract

Research in robotic ultrasound is steadily expanding. However, its introduction to clinical practice is still awaiting. One reason for this is the dependency on other imaging modalities for navigation. These approaches become cumbersome, especially in cases where other imaging data is unavailable. Furthermore, patients are unfamiliar with robotics in the healthcare domain and thus show low acceptance of these robotic systems.

Therefore, this thesis explores navigation approaches for robotic ultrasound scanning based solely on ultrasound images. Navigation approaches leveraging confidence maps are explored to improve the 3D compounding quality. Results are promising in reducing high-reflector-induced shadow areas in compounding. Furthermore, a navigation approach based on automatic live segmentation of thyroid tissue was evaluated in a phantom study. The robotic volumetry approach led to superior results compared to the clinical standard. Physics-inspired data augmentations were introduced to enable a more explainable training of ultrasound imaging networks. The simulation and thyroid phantom studies show promising results for ultrasound-guided robotic ultrasound scanning. The contributions of this thesis further pave the way toward fully autonomous robotic ultrasound scanning.

Lastly, this thesis touches on patient acceptance in medical robotic imaging. As a pioneering work, we showed that introducing interaction and communication in a robotic examination can increase acceptance and reduce stress. We evaluated these interactions in the case of robotic ultrasound scanning. However, the learnings can be extended to other medical robotic imaging systems, thereby increasing acceptance which can help bring robotic imaging systems into clinical practice.

Zusammenfassung

Die Forschung im Bereich des robotischen Ultraschalls nimmt stetig zu. Die Einführung in die klinische Praxis lässt jedoch noch auf sich warten. Ein Grund dafür ist die Abhängigkeit von anderen Bildgebungsmodalitäten für die Navigation. Diese Ansätze sind umständlich, insbesondere in Fällen, in denen keine anderen Bildgebungsdaten verfügbar sind. Darüber hinaus sind Patienten mit der Robotik im Gesundheitswesen nicht vertraut und zeigen daher eine geringe Akzeptanz für diese Robotersysteme.

Daher werden in dieser Arbeit Navigationsansätze für robotergestützte Ultraschalluntersuchungen untersucht, die ausschließlich auf Ultraschallbildern basieren. Es werden Navigationsansätze erforscht, die 'confidence maps' nutzen, um die Qualität des 3D Ultraschallvolumens zu verbessern. Die Ergebnisse sind vielversprechend in Bezug auf die Verringerung der durch hohe Reflektoren verursachten Schattenbereiche bei der volumetrischen Rekonstruktion. Darüber hinaus wurde ein Navigationsansatz, der auf der automatischen Live-Segmentierung von Schilddrüsengewebe basiert, in einer Phantomstudie evaluiert. Der robotergestützte Volumetrieansatz führte im Vergleich zum klinischen Standard zu überlegenen Ergebnissen. Physikalisch inspirierte Datenerweiterungen wurden eingeführt, um ein besser erklärbares Training von bildgebenden Ultraschallnetzwerken zu ermöglichen. Die Simulations- und Schilddrüsenphantomstudien zeigen vielversprechende Ergebnisse für das ultraschallgeführte robotergestützte Ultraschallscanning. Die Beiträge dieser Arbeit ebnen den Weg zum vollständig autonomen robotergestützten Ultraschallscannen.

Schließlich befasst sich diese Arbeit mit der Patientenakzeptanz in der medizinischen robotergestützten Bildgebung. In einer Pionierarbeit haben wir gezeigt, dass die Einführung von Interaktion und Kommunikation bei einer Roboteruntersuchung die Akzeptanz erhöhen und Stress reduzieren kann. Wir haben diese Interaktionen am Beispiel der robotergestützten Ultraschalluntersuchung untersucht. Die gewonnenen Erkenntnisse lassen sich jedoch auch auf andere medizinische Bildgebungsroboter übertragen, wodurch die Akzeptanz erhöht und die Einführung von Bildgebungsrobotern in die klinische Praxis gefördert werden kann.

Acknowledgments

First and foremost, I want to thank Nassir for allowing me to pursue a Ph.D. at CAMP. You allowed me to work on topics that interested me and allowed for radical and creative thinking. Next, I also have to thank Thomas and especially Ben. Without the input and discussion from this trio, I would not be where I am today.

I also have to thank the Garching administrative team, Sara, Amir, and Martina. Without you, life at the chair would have been a much bigger hassle. So thank you!

I also want to thank all my students. It was great working with you and learning from you.

To my old and new IFL crew: thank you for making IFL the place it is. It was a pleasure to work with you, and I wish all of you the best for your future. And to my IFL core: you know who you are. I am glad that colleagues turned into friends, and I wouldn't want to miss having you in my life.

Lastly, a big thank you goes to my family and friends outside work. Even though you didn't always know what I was doing, I received continuous support over the last few years. I couldn't have done it without you.

Contents

Part I

Introduction

Introduction

1

Contents

1.1 Motivation & Objectives

Ultrasound (US) [1] is a well-established imaging modality in clinical practice due to its advantages. It allows for real-time imaging, is non-invasive, does not cause any radiation, and is affordable compared to MRI [2] or CT [3] imaging. Therefore, US is often used in invasive and non-invasive diagnostic tasks as well as during intervention. These applications include, for example, thyroid and liver diagnostics, artery plague detection, and intraoperative brain tumor removal. However, ultrasound also has its drawbacks. Due to its manual handling, ultrasound shows a high operator dependency and low reproducibility [4]. These manual acquisitions can also lead to musculoskeletal pain [5, 6, 7] and fatigue [8] if executed constantly. Furthermore, US is view-dependant and inherently produces a lot of speckle noise and artifacts in the image. These latter drawbacks make it difficult to interpret the US images, requiring extensive expert knowledge both for acquiring and analyzing the data.

From the 1990's on [9], 3D US compoundings started to gain popularity. These compoundings are volumetric representations of US data and can be created through specialized 3D US probes or by tracking a 2D US probe and compounding the acquired images. In the case of mechanical 3D probes, a linear array is automatically moved by rotating, tilting, or translating, creating several 2D US images that are combined in a 3D reconstruction. These probes enable short reconstruction times; however, they are larger than conventional probes and require steady handling; otherwise, errors are introduced. Furthermore, the 3D probes are standalone products and have to be additionally purchased application-specific by clinics. Some research progresses on 2D arrays. However, these probes are only developed and evaluated in research and are not yet available in clinical practice. In contrast, 2D probes are already widely available in clinical practice and are less bulky and less expensive than their 3D counterpart. These probes can be tracked in several ways, including electromagnetic, optical, and mechanical, or robotic. Electromagnetic and optical tracking both consist of a source and a sensor or marker attached to the 2D probe, respectively. These tracking mechanisms require less space than mechanical ones but cannot allow metallic objects close by or require dealing with line-of-sight issues. In contrast, robotic tracking requires a more extensive setup but also enables a higher tracking accuracy [10].

But why do we need 3D US compoundings? The answer to this question is manifold. The human anatomy is a volumetric construct. Therefore, mapping this information into a 2D view is inherently complex. With 2D US, physicians have to manually map a 2D slice into a 3D structure. This procedure requires a high cognitive load, and at the same time, it will be nearly impossible for a second physician to reproduce the exact same 2D image. This operator dependence creates a high variability in the acquired data. Physicians execute scans differently and choose different 2D B-mode images for diagnostic purposes. In contrast, 3D US enables the preservation of 3D information, thereby reducing variability. At the same time, standardization and repeatability are increased because diagnostic tasks do not depend on several B-mode images but on a more general 3D volume. In recent years, network-based segmentation and classification tasks also became more prominent. These automatic approaches can further reduce variability and assist physicians in diagnostic and interventional tasks, such as organ volume estimation or needle guidance. The networks can be applied to 2D or 3D US data to achieve these automated tasks successfully. Here, segmentations in 3D lead to the same advantages as mentioned before for the 3D US case. A single segmentation on a 2D slice does not provide volumetric information. Oppositely, segmenting an organ on a tracked stack of 2D US images followed by a compounding of the segmentation or segmenting on the 3D US compounding directly enables a volumetric representation of the organ of interest. Furthermore, a 3D US representation provides spatial information from neighboring frames, which can help to improve the segmentation performance. Therefore, 3D US compoundings are necessary to reduce variability and increase the standardization and accuracy of automated image analysis. **So how can we intelligently improve the computation of 3D US compoundings?**

We have seen that good US compoundings are essential for intelligent robotic US approaches. However, the compounding itself will not replace a standard clinical procedure. This replacement can only be achieved if a robotic acquisition is combined with a second task. For example, combining a robotically acquired US volume with an automatic downstream task can enable efficient segmentations of different anatomical structures. However, the segmentation alone is not showing its full potential if not tailored to a clinical diagnostic or interventional task. **So which clinical procedure can be improved through intelligent, robotic systems to make these systems applicable?** And how can we compare the intelligent and applicable robotic system to the current clinical gold standard?

Once it is shown that intelligent and applicable robotic systems can improve clinical procedures, these systems should be translated into clinical practice. However, system performance is not the only measure of success in this transition anymore. Compared to the other research questions, humans must be taken into account, and human factors become crucial. Hospital staff and patients will interact with robotic systems in clinical practice. Therefore, if one of the groups does not accept modern robotic systems, these systems will simply not be integrated into the hospital. These interactions will look different for physicians and patients. While a doctor might perceive the system as a tool that enhances his or her task, patients might see the robotic system as a partial replacement of the human sonographer. Therefore, different aspects that can influence acceptance should be considered for both groups, and in this thesis, we will focus on the patient group. So lastly, we are asking: **How can we ensure that patients accept intelligent and applicable robotic systems?**

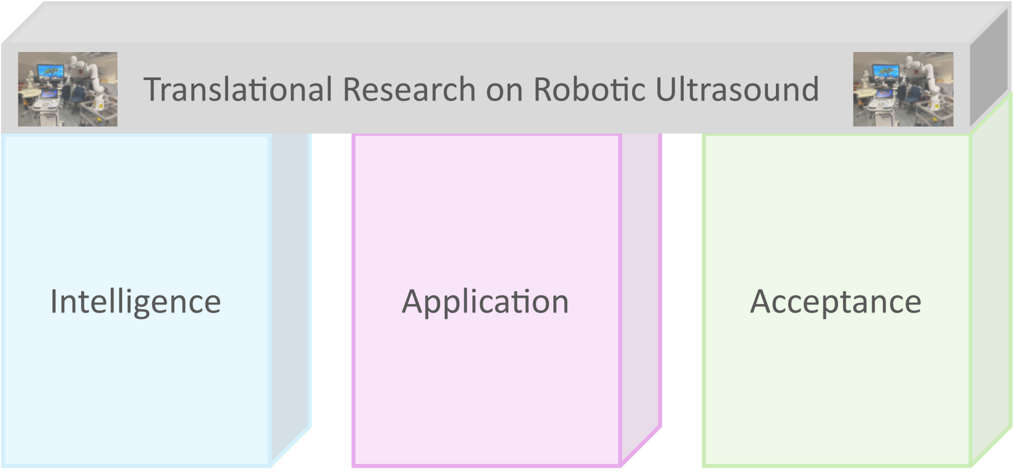


Fig. 1.1. The three pillars for successful translational research on robotic ultrasound: intelligence, application, and acceptance.

This thesis will try to answer all three questions and propose methods enabling intelligent, applicable, and accepted robotic ultrasound systems (see Figure 1.1).

1.2 Key Contributions

We will answer the questions from the previous section with the contributions described here.

The first two contributions enable intelligent robotic ultrasound by providing variable US data and improving US compoundings by selecting the best successive frames.

1. Creation of highly variable US data through physics-inspired augmentation.

> Maria Tirindelli*, Christine Eilers*, Walter Simson, Magdalini Paschali, Mohammad Farid Azampour, and Nassir Navab. "Rethinking Ultrasound Augmentation: A Physics-Inspired Approach". MICCAI 2021. [11] (*authors contributed equally)

We propose new data augmentation techniques for ultrasound images inspired by ultrasound physics. Currently, ultrasound data is augmented with techniques from the computer vision domain. These transformations, however, do not resemble the physical behavior of ultrasound imaging and therefore result in unrealistic ultrasound images. In this work, we, therefore, introduce augmentation techniques that mimic deformation, reverberation artifacts, and changes in signal-to-noise ratio. We evaluate these augmentations on a newly published dataset of spine US images.

2. Improvement of robotically acquired US compoundings for high-quality volumetric US data.

Viviana Sutedjo*, Maria Tirindelli*, Christine Eilers*, Walter Simson, Benjamin Busam, and Nassir Navab. "Acoustic Shadowing Aware Robotic Ultrasound: Lighting up the Dark". IEEE RA-L 2022. [12] (*authors contributed equally)

A modular framework for acquiring and improving robotic ultrasound compoundings is designed, implemented, and evaluated. In the first step, the robot efficiently covers the manually defined volume of interest. Afterward, acoustic shadowed areas are computed based on confidence maps. The robot then rescans these low-quality areas to improve the US compounding. Results show substantial improvements in the US compounding after applying the acoustic shadow reduction module.

The following contribution facilitates intelligent robotic US scanning for a specific clinical application:

3. Intelligent robotic sonography for accurate and repeatable thyroid volumetry.

John Zielke*, Christine Eilers*, Benjamin Busam, Wolfgang Weber, Nassir Navab, and Thomas Wendler. "RSV: Robotic Sonography for Thyroid Volumetry". IEEE RA-L 2022. ICRA 2022 presentation. [13] (*authors contributed equally)

We propose a robotic framework for automatic thyroid volumetry estimations. A robotic manipulator acquires US thyroid scans based on the live segmentation of thyroid tissue on US images. Once the robot scans both lobes, the thyroid issue is automatically segmented in the US images, the segmentation is compounded, and the thyroid volume is calculated. Results from a phantom study show significant improvements comparing the robotic approach and the current clinical 2D B-mode standard.

The final contribution targets patient acceptance of robotic US systems:

5. Increased patient acceptance through robot-human interaction and communication.

Christine Eilers*, Rob van Kemenade*, Benjamin Busam, and Nassir Navab. "On the Importance of Patient Acceptance for Medical Robotic Imaging". arXiv:2302.06208, 2023. [arXiv preprint] [14] (*authors contributed equally)

We analyze the effect of interaction and communication on acceptance in medical robotic imaging. To achieve this task, we propose an interaction pipeline for robotic ultrasound scanning of the neck. Additionally, we evaluate this pipeline with qualitative and quantitative measures. Questionnaire results indicate that robot-patient interaction and communication can improve the acceptance of robotic ultrasound systems.

1.3 Outline & Overview

Before answering the raised research questions, we will briefly summarize the manuscript here. While the initial chapters give a general overview, later chapters discuss the scientific contributions for enabling intelligent, applicable, and acceptable robotic ultrasound in the context of the current state of the art. The chapters are structured as follows:

Chapter 1 introduces and motivates the work while pointing out key contributions and providing an overview of the thesis.

Chapter 2 explains the fundamentals of ultrasound.

Chapter 3 provides the basics of robotic manipulators.

Chapter 4 gives a short overview of general robotic ultrasound.

Chapter 5 describes the main setup used throughout the thesis.

Chapter 6 introduces physics-inspired US augmentations to enable intelligent and explainable imaging tasks.

Chapter 7 proposes algorithms for intelligently improving robotically acquired ultrasound compoundings.

Chapter 8 illustrates the applicability of robotic ultrasound for the clinical application of thyroid volumetry.

Chapter 9 discusses patient acceptance and the influence of robot-patient interactions in robotic ultrasound scanning.

Chapter 10 critically summarizes the thesis and provides an outlook.

Before diving into the scientific contributions, we will provide fundamental knowledge of ultrasound, robotic manipulators, and robotic ultrasound in the next part.

Part II

Fundamentals

In Part II of this thesis, we will provide fundamental knowledge on relevant research areas for this work. First, we will look at ultrasound imaging, including the physics of ultrasound, its appearances and acquisitions, and finally, volumetric ultrasound compounding. Next, we describe the basics of robotic manipulation, divided into design, control, and autonomy elements. Last, we will give a brief overview of robotic ultrasound in general.

Ultrasound

2

Contents

Ultrasound (US) is a well-established medical imaging modality. Due to its advantages, ultrasound is widely used in clinical practice, especially in diagnostic tasks. Its real-time capabilities allow for easy manipulation, quickly viewing different areas of interest from different views, and tissue analysis through deformation. Furthermore, ultrasound is radiation-free and more affordable than computational tomography (CT) and magnetic resonance imaging (MRI). However, ultrasound also has its drawbacks. The physics of ultrasound lead to a significant amount of speckle noise, making the analysis of US images difficult. Furthermore, the probe has to be moved manually by an operator. Finding the right view, correct US settings, and analyzing the images correctly, therefore, requires a lot of training and experience. This manual operation also introduces a high operator dependency such that acquired images for the same task can vary, and results are not always comparable.

2.1 Physics of Ultrasound

In ultrasound imaging, sound waves create an image of the underlying anatomy. Due to different tissue characteristics, sound waves are absorbed and reflected differently per tissue, forming an image of the viewed anatomy. Similar to other types of waves, US sound waves are characterized by their **frequency** and wavelength. The frequency describes the number of cycles per second, while the wavelength describes the length over which one cycle occurs. Both parameters are inversely related. Thus a high frequency leads to a short wavelength and vice versa. In ultrasound imaging, it is, therefore, crucial to decide on the correct frequency for specific applications. A high frequency leads to a high resolution in superficial areas, but deeper anatomies are not visualized clearly. In contrast, a lower frequency enables the penetration of deeper tissue levels.

Ultrasound is prone to **attenuation**, which describes the loss of amplitude and intensity. This behavior is affected by frequency, depth, probe angle, and tissue characteristics. A high frequency, for example, leads to rapid attenuation. Attenuation can occur in four ways: absorption, reflection, refraction, and scattering (Fig. 2.1). During **absorption**, the sound wave is converted to heat, thereby reducing the amount and intensity of continued or reflected sound waves. **Reflection** describes the process of sound waves being reflected back to the

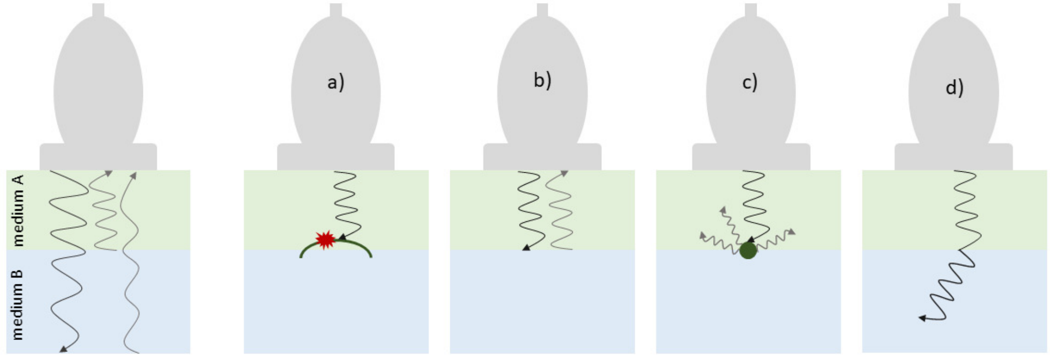


Fig. 2.1. US attenuation mechanisms: a) absorption, b) reflection, c) scattering, d) refraction.

transducer. Most reflections occur when the sound wave hits the medium perpendicularly. In non-perpendicular angles, fewer and weaker sound waves are reflected. During **refraction**, the sound wave bends on the border of two different tissue layers at an oblique angle, changing the direction of the original sound wave. Lastly, **scattering** occurs when the sound wave hits a medium smaller or equal to the size of the sound wave. The sound wave then gets reflected in many directions, resulting in only a few echoes reaching the transducer.

2.2 Characteristic Ultrasound Appearances

Different tissues show characteristic appearances on the ultrasound image based on the interaction between tissues and sound waves. In general, the density, smoothness, stiffness, and surface size of the tissue influence the amount and strength of the reflected signal. Fluid-filled structures, like arteries and veins, are **anechoic** and appear black. In contrast, strong reflectors, like bones, are **hyperechoic** and therefore appear in bright white. Furthermore, these highly reflective structures introduce **acoustic shadows** behind them because little to no signal penetrates the strong reflector. Solid organs, like the liver, kidney, or thyroid, appear grey. Gas scatters the sound waves, creating small hyperechoic foci.

These specific ultrasound characteristics also introduce artifacts in the images (Fig. 2.2). Here we will name and explain a few of them. **Acoustic shadows** appear behind highly reflective structures. These structures, for example, bones, entirely absorb or reflect the incoming sound waves such that no sound wave reaches behind the structures. As a result, the US image behind these reflective structures will be completely black. Similarly, **edge shadowing** appears as thin acoustic shadows behind the lateral edges of cystic structures. This phenomenon exists because sound waves hitting a structure at a tangential angle are refracted. **Reverberations** occur due to the bouncing of sound waves between two highly reflective structures. The US system detects the bouncing as a prolonged travel time and correlates it to a longer distance between the transducer and the anatomical structure. This prolonged echo then results in a copy of the reflective structure in the US image deeper than the original structure. **Comet tails** are similar to reverberations, showing additional bright lines in distance to a highly reflective structure due to a front and back between non-parallel and strongly reflective structures.

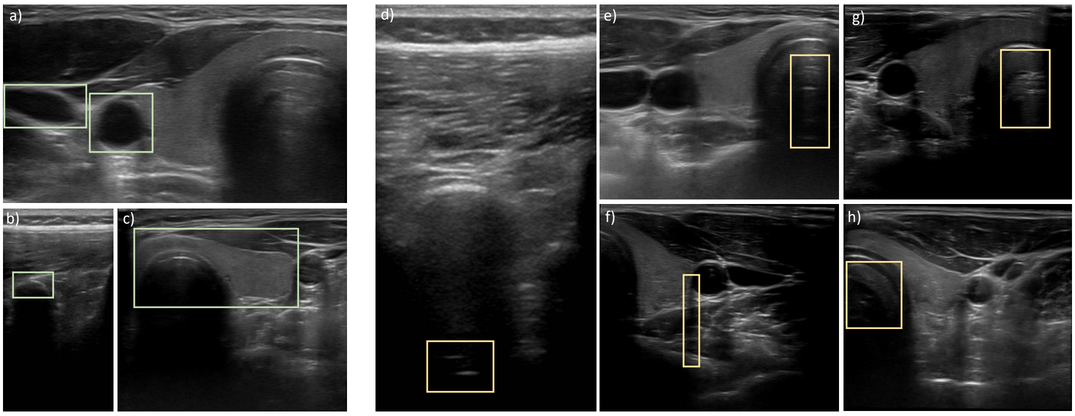


Fig. 2.2. US image and artifact examples: a) arteries and (compressed) veins appear black, b) bone as a highly reflective structure appear bright, c) thyroid as a solid organ in grey, d) and e) reverberation artifacts, f) edge shadow, g) and h) mirroring. All images show acoustic shadows below the bones and the trachea surface.

Lastly, in **mirroring**, parts of the US image are duplicated on the opposite side of a strong reflector.

2.3 Ultrasound Machines and Acquisitions

In ultrasound transducers, sound waves are generated through piezoelectric elements. On the one hand, crystals inside the transducer vibrate when voltage is applied, creating a sound wave. On the other hand, the crystals vibrate when sound waves hit the transducer and create an electric current which is then transformed into an ultrasound image. In this way, the transducer simultaneously works as a sender and receiver of sound waves. **Beamformer** in the US probe steer and focus the transducer elements on the probe to achieve a desired image. Thereby, this part of the US system computes the correct pattern for the transducer array to send and receive sound waves. Improving beamforming is a research topic on its own and will not be discussed in this thesis. However, further general information and current research directions can be found here: [15, 16].

Transducers for ultrasound imaging differ in shape and application. A linear probe has a rectangular field of view, which enables high-frequency imaging. Therefore, it is mainly used for vascular and superficial structures, such as arteries, veins, and the thyroid. Convex and phased array probes both work with low frequencies. The convex probe enables a trapezoid field of view and is mainly used in abdominal scans, such as the aorta or liver scanning. In contrast, the phased-array probe creates a sector-shaped field of view and is generally used in cardiac applications.

The ultrasound frames can then be acquired in three different planes with regard to the patient's body. The **transverse** plane separates inferior from superior, meaning the top of the patient from the bottom of the patient. The **saggital** plane separates left from right, and the **coronal** plane separates the front from back.

2.4 Volumetric Ultrasound Compounding

Through **volumetric ultrasound compounding**, ultrasound can be elevated to a 3D format (see Figure 2.3. By tracking the position of the US probe, 2D images are stacked into a 3D volume. Therefore, US compounding is an extensive research area with many advances regarding data acquisition and compounding algorithms.

Data for 3D US can be acquired in different ways. One way is through the use of 3D wobbler probes. In these probes, an array of piezoelectric elements automatically performs a small sweep in a given space. However, these probes are larger and more expensive than conventional 2D US probes. Furthermore, they show a lower resolution and only have a small volumetric field of view. Another option for generating volumetric compoundings is the tracking of 2D US probes. This tracking can be achieved in different ways, including electromagnetic, optical, acoustic, mechanic, and inertial measurement unit (IMU)-based tracking. All options have different accuracies, advantages, and drawbacks and therefore have to be chosen based on the specific application. Electromagnetic tracking, for example, is sensitive to metallic objects. Therefore it can be used for diagnostic tasks but is not favorable in an interventional setting. Optical tracking is prone to line-of-sight issues and, therefore, highly dependent on the hardware setup and people's movement in clinical rooms. Mechanical or robotic tracking requires more extensive hardware on site but also allows for the highest accuracy [10]. Lastly, approaches are developed to track US images without external sensor information. These deep-learning-based approaches use IMU and speckle information, such as linear regression and decorrelation, to gather spatial information about US frames [17, 18].

After the data acquisition, the frames are reconstructed into a 3D volume to translate pixel information into voxel information. Most approaches use a combination of interpolation and averaging to achieve this task. The approaches can further be divided into four groups: voxel-based, pixel-based, function-based, and hybrid versions. In the **voxel-based** approach, the algorithm loops through all voxels in a target volume and assigns corresponding pixels from the input images to each voxel. The voxel value is determined based on one or more pixel values. If more than one pixel is used, the pixel values are usually interpolated to compute the voxel value. The **pixel-based** approach is a reversed method. Here the algorithm loops through all pixels of the input images and assigns them to one or more corresponding voxels in the target volume. Most approaches in this category consist of two steps: distribution and hole filling. In the first step, pixel values are assigned to one or several voxels. However, this can lead to empty voxels. So in the second step, these voxels are filled with reasonable values. Lastly, in **function-based** methods, one estimates functions based on the input data that are later used to create the voxel grid. These functions are often several polynomials, and the coefficients are determined such that the functions pass through the input voxels. Then, a regular voxel array is created by evaluating the functions at regular intervals.

We already saw that robotic tracking enables the most accurate tracking results. So next, we will look at the basics of robotic manipulation, specifically the design, control and autonomy levels of these robot types. Understanding both US imaging and robotic manipulation then allows us to combine both areas into high-quality robotic ultrasound.

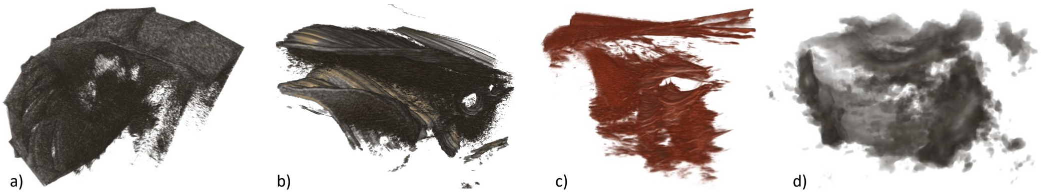


Fig. 2.3. Volumetric US compounding examples, all from internal datasets: a) simulated abdomen, b) thyroid phantom, c) healthy volunteer thyroid, d) transrectal patient US.

Robotic Manipulator

3

Contents

Robots can appear in many different shapes and control types, such as humanoid robots, legged robots, or **robotic manipulators**. This work focuses on the latter ones, also called robotic arms. Similar to all other robot types, these can be programmed to perform a specific task autonomously and intelligently. Robotic arms were first introduced in the industrial area but are also becoming prominent in the medical field since the 1980s. In 1985 a robot was used for CT-guided brain tumor biopsy [19], marking the beginning of medical robotics. Since then, robotic manipulators have been adapted for the medical field and used in different areas such as surgery, diagnostics, and rehabilitation. Furthermore, these robots have been combined with different clinical tasks, such as imaging, biopsy, or surgical tasks.

3.1 Design

A usual robotic manipulator consists of rigid links connected by rotational joints. One side of the robotic manipulator is fixed at a base, while the other side, the **end-effector**, can be equipped with application-specific tools to perform a desired task. These tools can be, for example, surgical tool holders, needles, microscopes, or ultrasound probes. Most robotic manipulators consist of six to seven joints because this design enables a range of motions similar to a human arm. These robotic arms usually have six degrees of freedom ($m_0 = 6$), meaning the end-effector can be manipulated in three translational and rotational directions. A seven-joint robotic manipulator ($n = 7$) shows an additional advantage called **redundancy**. In this case, the amount of joints exceeds the degrees of freedom ($n > m_0$), allowing for several joint value solutions for a given end-effector pose. This property allows all other robotic parts to be moved while the end-effector stays in place. In the medical field, this feature can be beneficial in several tasks. For example, during US-guided needle biopsies, the US probe can stay in a pre-defined position, while the remaining parts of the robot can be moved based on the preferences of the clinical staff.

The joints of robotic manipulators generally contain torque sensors. These sensors allow us to measure torques at the joints and therefore allow us to infer applied forces along the robotic arm and at the end-effector. These forces arise from the robot's movement and contact with objects or humans. In some manipulators, an external force sensor is also attached to the end-effector, allowing one to measure the forces applied directly. Therefore, these additional

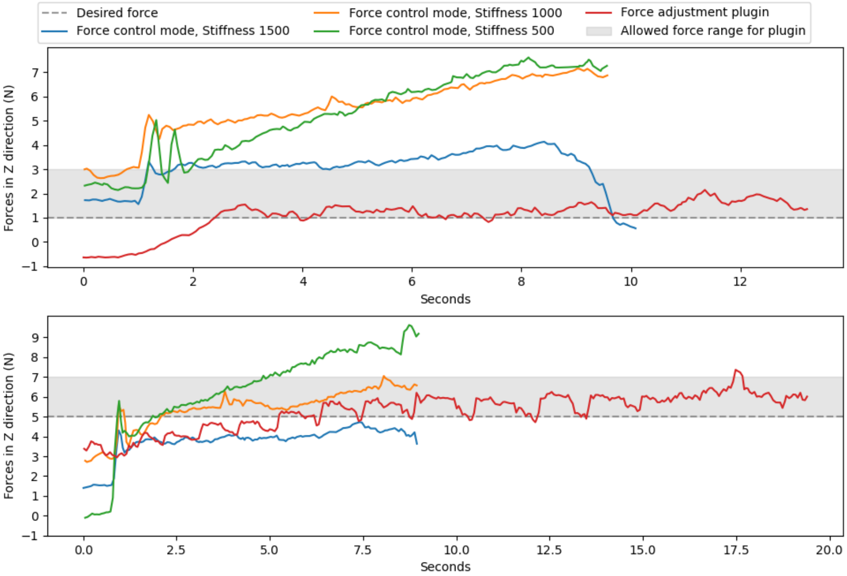


Fig. 3.1. Force measurements for four different force controls during thyroid scanning. The force control mode describes a force control based on the force estimations at the end-effector from torque sensors in each joint. The force adjustment plugin (red) replaces the information from the torque sensors with an external force sensor at the end-effector. It can be seen that the latter method is significantly more accurate and stable in keeping the desired force. Force control mode was enabled through iiwa stack [20]. Figure from [21].

external force sensors are more accurate in measuring forces applied to the end-effector itself. Figure 3.1 shows an example of this behavior in the case of thyroid US scanning, comparing force control through torque sensors with force control through an external force sensor.

In this work, we will not analyze the design and modeling of robotic arms in more detail. However, if the reader is interested in this topic, we suggest the following resources: [22, 23, 24].

3.2 Control

The work presented in this thesis is based on fundamental knowledge of robotics and kinematics. For more general and basic knowledge, we, therefore, refer to the following sources: [25, 26]. Robotic manipulator control types can follow forward or backward kinematics and can be based on position control, force control, or a hybrid version. In the following, we will have a brief look at all of these types.

The end-effector position can be described in cartesian and joint space. In joint space, the end-effector position is therefore defined by the value of each joint, such that $P[\theta_0, ..., \theta_n]$. In contrast, in cartesian space, the end-effector position is given by a $4x4$ matrix consisting of orientation and position such that

$$P = \begin{bmatrix} R & p \\ 0 & 1 \end{bmatrix}$$

with R being a 3×3 rotation matrix and p being a 3×1 position vector, both described in relation to a base coordinate system which is usually located in the robot base.

Then in **forward kinematics**, the joint values are given to compute the end-effector pose either through trigonometry, Denavit-Hartenberg parameters, or the screw theory. Vice versa, in **inverse kinematics**, the end-effector pose is given, and the corresponding joint positions have to be computed. This problem can be solved through analytical or numerical approaches, whereby the pseudoinverse Jacobian method is frequently used for the latter case. Furthermore, inverse kinematics is used more regularly in robotic control because the cartesian workspace of the robot is more intuitive to users than its joint space. However, inverse kinematics also requires a slightly longer computation time. Therefore, applying the joint space for stopping criteria or other time-critical tasks can be advantageous.

Initially, robot control schemes can be divided into **position control** and **Force control**. In a position control scheme, like a PID control, the desired and current joint positions are given, and an error between both states is computed to move the robot accordingly. This approach is effective but rarely considers the robot's dynamics and is, therefore, rarely used on its own. On the other hand, force control adapts the force executed by the robot and can be executed directly or indirectly. In the direct case, an estimated state and the desired force are given, and the required force is computed. However, this approach can lead to undesired behaviors if the robotic manipulator is not touching the object yet. In that case, no counterforce is present, and the robot accelerates to achieve the desired force, leading to fast and uncontrollable movements.

To circumvent this problem, **impedance control** can be applied. With this approach, we can model dynamic behaviors between the robotic end-effector and the environment. The interaction between both parts is modeled as a virtual mass-spring-damper system and described through damping and stiffness. While a high stiffness nearly permits motion in the direction of the virtual spring, a low stiffness enables a more extensive motion range around the desired position. The adequate chosen stiffness then also depends on the environment. A stiffer environment should result in a softer model stiffness exerted by the robot and vice versa. In most robotic applications, the end-effector only exerts a contact force in one direction. In this case, a soft motion control is enabled in the direction of contact, while a stiffer motion control is executed in other movement directions. Impedance control only indirectly assigns contact forces by controlling the position. The forces are, therefore, also only estimations from the joint torque values. An impedance control scheme can be combined with external force sensing for more accurate force control. Many robotic manipulator control schemes apply a hybrid impedance and position control version. While impedance control exerts a steady force onto an object, position control enables the robot to follow a trajectory on the two other axes.

Now that both basics of ultrasound and robotic manipulators have been explained, we will analyze their combination, resulting in robotic ultrasound, in the next chapter.

3.3 Levels of Autonomy

Robots and robotic manipulators can be grouped into different classes, depending on their amount of autonomy. These categories range from manual to full autonomy. Beer et al. [27] introduced a framework for the level of robotic autonomy (LORA) in human-robot interaction and grouped classes based on autonomy and the amount and type of tasks executed by the robot and the human (see Table 3.1).

LORA	**Sense**	**Plan**	**Act**
Manual	Human	Human	Human
Teleoperation	Human/Robot	Human	Human/Robot
Assisted Teleoperation	Human/Robot	Human	Human/Robot
Batch Processing	Human/Robot	Human	Robot
Decision Support	Human/Robot	Human/Robot	Robot
Shared Control Human Initiative	Human/Robot	Human/Robot	Robot
Shared Control Robot Initiative	Human/Robot	Human/Robot	Robot
Executive Control	Robot	Human	Robot
Supervisory Control	Human/Robot	Robot	Robot
Full Autonomy	Robot	Robot	Robot

Tab. 3.1. Different levels of robot autonomy with human and robot involvement in the sensing, planning, and acting parts. Adapted from [27].

Yang et al. [28] defined the levels of anatomy in the context of surgical robotics and introduced the following groups, ranging from least to most autonomy:

- No autonomy: the operator performs all cognitive and executive tasks.
- Robot assistance: the operator remains in control, and the robot provides assistance.
- Task autonomy: the operator generally stays in control, but specific tasks are executed by the robot automatically after being initialized by the operator.
- Conditional autonomy: the operator selects, approves, and monitors the surgical task while the robot automatically performs the chosen procedure.
- High autonomy: the robot can make decisions and perform the procedure alone, but the operator is supervising it.
- Full autonomy: the robot performs the procedure automatically, and no human is involved.

Most research is done between robot assistance and high autonomy. Medical robotics can also achieve full autonomy on a technological level. However, regulatory, ethical, and legal aspects have to be analyzed and discussed in parallel before enabling the introduction of fully autonomous medical systems.

Robotic Ultrasound

4

Contents

Robotic ultrasound (RUS) combines the advantages of ultrasound and robotics to create an updated version of ultrasound imaging. Up until today, several review papers have been published on the topic of robotic ultrasound. We refer the reader to these reviews for a more detailed analysis [29, 30, 31, 32]. Robotic ultrasound systems (RUSS) can be classified based on their autonomy level and use case. This chapter will focus on extracorporeal diagnostic systems and briefly overview the other clinical areas. We do not provide completeness for all areas but give a small insight into different developments in robotic ultrasound.

4.1 Extracorporeal Diagnostic

Contrary to initial thoughts, researchers did not develop systems with increasing autonomy after another. In contrast, up until today, teleoperated, shared control, and fully autonomous systems have been designed more or less in parallel. One of the lowest levels of autonomy is described as telemanipulation. In this case, a sonographer remotely manipulates a robot through a joystick to perform a US examination [33, 34, 35, 36]. In a teach-and-replay system, the sonographer manually moves the US probe, and the robot trajectory is recorded. Then the robot can automatically execute the US scans based on the prior demonstration [37, 38, 39, 40]. The next level of autonomy shows human-robot collaboration. In these scenarios, the robot is still telemanipulated, but visual servoing techniques are applied, for example, to keep the area of interest in the image center. Early works of this shared control started in 1999 with a 6 DoF robotic arm from Salcudean et al. [41, 42]. This work for carotid artery examinations used a shared control strategy. While the physician remotely controlled the US probe movement along the vessel, the robot adjusted its motion in the two other directions to keep the artery in the center of the image. Abolmaesumi et al. also developed US image-based servoing for shared control early on [43, 44, 45]. Further research on shared control includes virtual fixtures [46] and dual force sensing [47]. Arbeille et al. also incorporated force feedback into telemanipulation and exchanged the until then standard joystick controller with a copy of the US probe on the master site [48, 49]. Many research projects are getting close to being fully automatic. While the robot can execute the scan itself automatically, the probe has to be placed initially by an operator [50, 51, 52]. Gourdon et al. [53] introduced a cage-like probe holder for US scanning. This device could scan the area it was placed on automatically but had to be held in place by an assistant.

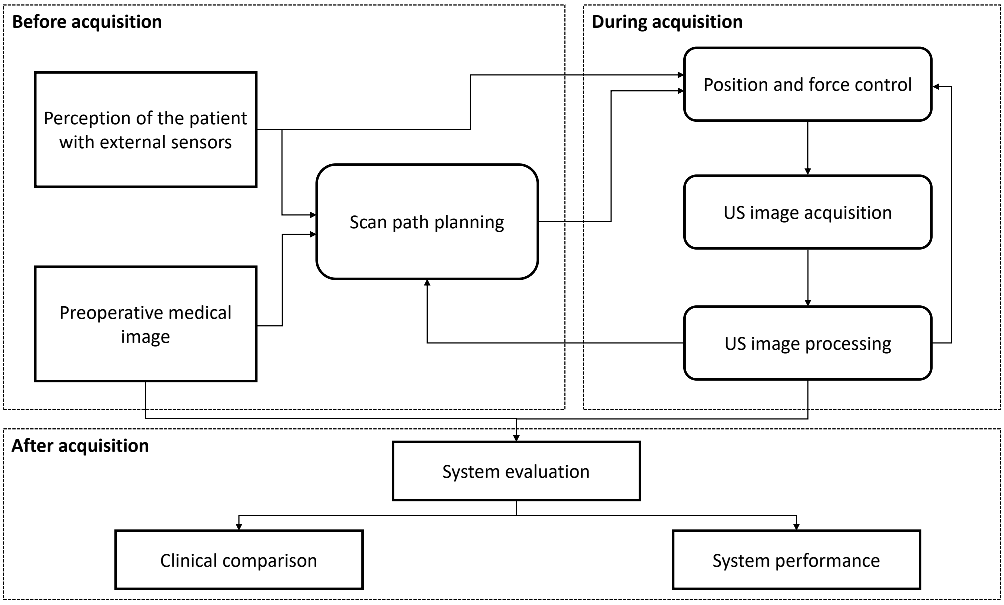


Fig. 4.1. Overview of the workflow of autonomous robotic US systems. Before the acquisition, preoperative data can be taken, and the patient can be scanned with external sensors. Both approaches then enable path planning. During the acquisition, position and force control is applied, and the US images are acquired and analyzed. Based on external and US data, the position and force control can then be adapted in real time. Lastly, after the scan is acquired, the system is evaluated in one of two ways: either by comparing it to a clinical standard or by evaluating the system's performance using statistical methods and metrics. Adapted from [32].

Next to telemanipulated and shared control, researchers developed robotic US systems and autonomous system types as well. These systems generally follow a specific workflow (see Fig. 4.1). This workflow consists of three building blocks: tasks before, during, and after the acquisition. Before the acquisition, the system uses preoperative data and patient reconstructions from an external sensor for path planning. These planned paths are then executed through position and force control while simultaneously acquiring US images. The US data can then be used to update the overall scanning path and to adjust the position and force control in real time. Furthermore, post-processing of the acquired US data is used for system evaluation after the acquisition. The systems can either be evaluated based on a clinical task or the system's performance, such as registration or segmentation accuracy.

Extensive work on visual servoing has been performed by Krupa et al., using image-region tracking to keep a target in the center of the US image [50], including speckle information in the servoing task [54] and extending moment-based visual tracking to the analysis of three orthogonal planes by Nadeau and Krupa [55, 56]. Working towards autonomous systems, Nakadate et al. automated carotid artery examinations by proposing completely automatic scanning methods [57, 58]. Mustafa et al. [59] used a commercial robotic manipulator for automatic liver scanning. A camera located the epigastric region of the patient, and the autonomous scan mimicked a manual US liver examination. Contrary to robotic manipulators, some research works combine US scanning with a robotic cage-like system attached to the patient bed [60, 61]. Welleweerd et al. [62] showed an automatic US screening of the breast based on a pre-defined trajectory and real-time visual servoing for probe adjustment. Other works include 3D US acquisition for improved diagnosis and repeatability. Merouche et al. [63] applied this concept to scan the lower limb arteries autonomously and create a 3D reconstruction to assess the length and severity of stenosis. In other works [64, 65], the scanning path is defined on preoperative MRI data. Scanning the patient with an RGB-D camera allows the MRI to be registered to the patient, and the robot automatically follows the planned trajectory. Similarly, with information from preoperative data, the robot can also scan regions of interest instead of following a pre-defined trajectory [66]. Göbl et al. [67] further advanced these systems by planning to cover a region of interest on preoperative CT but also taking acoustic windows into account. Apart from preoperative data, the scanning path can also be defined manually or automatically on patient point clouds from RGB-D data [4, 68, 69]. Lastly, planning can be done on an atlas, as shown in work by Kaminski et al. [70] for thyroid scanning. The template is matched to the output of the RGB-D camera, and the robot can execute the preplanned with constant force. Von Haxthausen et al. [71] proposed a system for autonomous screening of peripheral arteries that automatically follows these arteries through a deep-learning-based vessel detection. Jiang et al. introduced work on motion- [72] and deformation-aware [73] robotic US. Furthermore, research recently has bee conducted on combining robotic US scanning with reinforcement learning [74, 75].

4.2 Needle Guidance

Many clinical tasks, such as biopsies and brachytherapy, require accurate needle placement. Ultrasound is often for guidance. However, manual positioning of the probe requires operator skill and very steady maneuvering. Furthermore, the US probe and needle often have to be coordinated simultaneously by one clinical expert, which requires a heavy load of hand-eye coordination. Robotic US guidance systems have therefore been developed to reduce or eliminate these drawbacks. In the developed techniques, needle insertion, US guidance, or both parts are performed robotically. Megali et al. [76] proposed an initial work for RUS-guided biopsy. The robot acquires a 3D volume of the area of interest. Then target point and point of entry are defined on the 3D volume, and the robot automatically aligns itself such that the needle can be inserted along the line between the entry and target point. Amongst others, Kettenbach et al. [77], and Pua et al. [78] worked on an automatic needle insertion while US served as the guidance method. Mallapragada et al. [79] introduced a robotic US system for breast biopsies in which the US probe repositioned itself to counteract target motion. The task of improving breast biopsies was also tackled by Nelson et al. [80], and Liang et al. [81] with both performing automatic US scanning and needle insertion. Several works focus on applying robotic US to prostate biopsies as well [82, 83, 84]. During brachytherapy, radioactive sources are placed inside the body to treat cancer. Robotically automizing this procedure has been steadily evolving. Starting with the first works on fully robotic brachytherapy by Ng et al. [85] and leading to more sophisticated systems, such as those proposed by Yu and Podder and colleagues [86]. Kojcev et al. used two robotic arms for US-guided needle placement [87]. One robotic arm is used for the autonomous acquisition of a 3D US volume and visual servoing during needle insertion, while the other arm inserts the needle. Li et al. [88] proposed a robotic system for needle insertion and guidance in liver tumor ablation. Chen et al. [89] introduced a system combining near-infrared and ultrasound for vessel imaging, assisting in robotic venipunctures.

4.3 Intraoperative

RUSS systems have also been developed for several intraoperative scenarios. In 1998 Cunha et al. introduced US-guided telesurgery for laparoscopic US imaging [90]. Here, the physician remotely controlled the US probe through a joystick. Research in this area has further progressed with the development of specific US guidance tools for the daVinci system, for example, by Budde et al. [91]. Additional works show US tool development directly interfacing with the da Vinci system. These tools include US probes on robotically actuated da Vinci tools [92], pick-up transducers [93], and a combination of robotic US imaging and elastography [94]. Ng et al. [85] and Mei et al. [95] developed a robotic US system for minimally invasive prostate ablation. In the first version, the robotic system automatically performs preoperative US prostate scanning. In the next version, this system also allowed for semi-automatic prostate ablation. In orthopedic surgery, research has been conducted to use robotic ultrasound for planning and implant beds in lateral skull bases [96] and bone cement detection in revision total hip displacement [97]. Zhan et al. [98] proposed a robotic US system integrated into the da Vinci setup. The US scanning is defined preoperatively but can be adjusted based on real-time tissue motion estimations. Langsch et al. [99] proposed a system consisting of an LBR iiwa and a 2D probe for autonomous catheter tracking during endovascular aneurysm repair. Esteban et al. [100] performed the first clinical trial for US-guided facet join insertion. Planning and insertion are done on US data, thereby eliminating the need for X-ray imaging. Torres et al. [101] developed a robotic system for navigation assistance during femur drilling.

As we saw in this small review and will see throughout the thesis, the advantages of robotic US are manifold. A high-accuracy tracking of the probe enables the creation of a volumetric 3D compounding of the US data. Furthermore, using robotics helps to apply a constant force on the tissue and repeatable scans. Lastly, RUS can decrease musculoskeletal pain for physicians who currently perform manual US scans during most of the day. These automatic approaches can also improve care in rural areas because robotic scans can be performed without a physician.

Next, we will look at the general setup used throughout the thesis.

Part III

Setup

Setup

5

Contents

The setup throughout the contributions presented in this thesis will be fairly similar. Therefore, we are dedicating a small chapter to the setup itself. This setup generally consists of three parts: the robotic system, the ultrasound system, and additional software. These components will be described briefly in the following sections.

5.1 Robotic System

All presented robotic works are implemented on a **KUKA LBR iiwa 7 R800** (KUKA AG, Augsburg, Germany) (Fig. 5.1, left), a robotic manipulator for human-robot interaction. It consists of 7 rotational joints, can handle a load of 7 kilograms, and has a reach of 800 mm. The provided software allows for both manual and automatic control. Manual control, enabled in joint or cartesian space, is executed through the provided SmartPad, a tablet that enables manual control and shows robot state information. The system also enables hand guidance of the robot. In this mode, the robot is maneuvered by moving it to its desired position by hand. The hand guidance mode can be used in combination with the automatic mode. In the automatic case, the robot runs robot applications that users can implement in the KUKA Sunrise API.

KUKA now also offers a medical-certified robotic manipulator, the **LBR Med**. This robotic arm and its control structure are very similar to the LBR iiwa setup. Therefore, the work presented in this thesis can be easily transferred to the medical-certified robot version and be applied in the clinical context.

The robotic manipulator is controlled through the KUKA Sunrise API based on Java. Controlling the robot this way is cumbersome, especially if combined with medical or optical data. Colleagues from the chair, therefore, developed **iiwa stack**[1] [20] (Fig. 5.1, middle). This meta-package connects the KUKA Sunrise workbench and the ROS environment. Thereby, this package allows programming the robotic manipulator through ROS via C++ or python code. It requires running a robot application on the robot side, enabling communication with and controlling the robot through a different workstation.

---

[1] `https://github.com/IFL-CAMP/iiwa_stack`, visited on: 14.02.2023

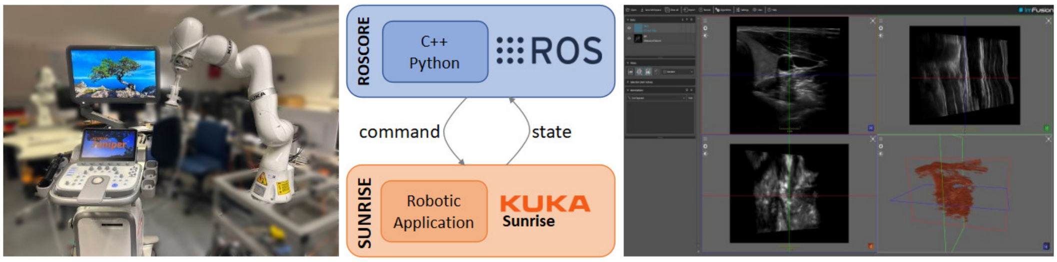


Fig. 5.1. Components of the general setup from left to right: US machine (Acuson Juniper) and robotic manipulator (KUKA LBR iiwa), communication scheme through iiwa stack (adapted from[2]), and view of the ImFusionSuite.

The end-effector of the robotic manipulator allows for attaching different tools. In general, we attach a US probe. In addition, we can add external force sensors for accurate and redundant force measurements and optical cameras to reconstruct the patient surface, for example.

5.2 Ultrasound Machine

In the presented works, we use a **Siemens Acuson NX3** ultrasound machine (Siemens Healthineers, Erlangen, Germany) (Fig. 5.1, second from left). This machine is commercially available and used for image-based diagnostic tasks in clinical practice. Next to B-mode imaging, the machine also provides additional information, such as Doppler imaging. Commercial systems like the one mentioned here own superior image quality. However, they do not provide direct access to the US data. Therefore, we are using a frame grabber to grab the US images from the US device screen. In this way, the US screen is connected to a workstation through a USB interface. Furthermore, we designed custom 3D-printed holders that carry linear and convex US probes and can be attached to the robotic arm rigidly.

5.3 Additional Software

ImFusion[1] (Munich, Germany) is a consulting and research and development company focusing on medical image computing and computer vision. Their software, **ImFusionSuite** (Fig. 5.1, right), allows for the intuitive handling of medical data and is therefore used in some parts of this thesis. Amongst others, medical imaging data can be visualized, annotated, and segmented. Furthermore, the academic version enables the creation of our own plugins for robotic control and medical image processing. Within the presented thesis, we will use this feature frequently and refer to some built-in algorithms provided by ImFusion.

In the next part, we will look at the optimization of US data, both for the data itself and tailored to 3D US compoundings.

---

[1] `https://www.imfusion.com/`, visited on: 14.02.2023

[2] `https://github.com/IFL-CAMP/iiwa_stack/wiki`, visited on: 14.02.2023

Part IV

Robotic and Ultrasound Optimization

So how can we intelligently improve the computation of 3D US compoundings?

In this part, we will answer this question through two different approaches. First, we will analyze ways to increase the size of US data sets. After investigating the current state of the art, we propose physics-inspired data augmentations, specifically for US imaging, to increase variability in US data sets accurately. After this, we will switch to US compounding optimization. First, we will look at current techniques to improve US compoundings. Then, based on gaps in the recent research, we propose to optimize US compoundings by mimicking the scanning approach of a human sonographer.

Ultrasound Data

6

Contents

6.1 Motivation

Ultrasound images can show a high variability due to differences in patient anatomies, US machine parameters, and acquisition parameters. Some of these attributes include but are not limited to phase aberration, scattering, and reverberation. Due to these artifacts, interpreting US data is challenging, especially for novice physicians and sonographers. Deep Neural Networks (DNNs) can assist image interpretation through tasks such as segmentation and registration, for example, for bone analysis [102, 103, 104, 105]. This enhanced interpretation could further improve diagnostic support and even robotic scans [106, 100, 74]. However, training DNNs requires large, diverse datasets, which are limited in medical US due to difficulties in acquiring and annotating data. Furthermore, these datasets should include a wide range of variability from factors such as operator dependence, US machines, acquisition parameters, and subject anatomies. This way, variable data could ensure the generalization of DNNs, and researchers could translate diagnostic tasks involving DNNs into clinical practice. Data augmentation depicts one method of synthetically increasing datasets. However, approaching this task to create more extensive and interpretable datasets is not straightforward, especially in the case of medical US. To tackle this problem, we introduce US-specific data augmentation techniques.

6.2 Related Work

Data augmentation, originating from the computer vision domain, is frequently applied to counteract limited training data [107]. DNNs trained on natural images can benefit from artificially enlarged datasets through different transformation techniques. These techniques can broadly be divided into two categories: synthetic data generation and image modification. Generative Adversarial Networks (GANs) [108] are used in the former category to produce synthetic data based on the available dataset. For instance, Zaman et al. [109] applied GANs to US images to generate data with varying time gain compensations, depths, and roll and pitch motions of the US probe. However, synthetic augmentations require substantial amounts of data to train the GAN itself. Therefore, these techniques become unsuitable when data is already very limited.

In the second category of data augmentation, classical image modifications are employed to increase the size and variety of the dataset. This augmentation method involves applying transformations such as random scaling, translations, rotations, and adding Gaussian noise to the images. Apart from the computer vision domain, various works applied these transformations to US data. Baka et al. [110] utilized mirroring and small free-form deformations on US spine images to improve bone segmentation. Other works have also applied random shifting [111, 112, 113], flipping [111, 114, 115, 116, 112, 117, 113], different translations [115, 118], rotation transformations [116, 102, 118, 117, 113], and varying brightness [114, 116].

However, these traditional augmentations are based on mechanisms from optical cameras, such as rotating the camera or zooming in. In contrast, US physics and image acquisitions fundamentally differ from natural images. Therefore, applying transformations from the computer vision domain to US data often results in unrealistic images (see Fig. 6.1). These newly created images increase the dataset but do not model the true variability of US acquisitions. A horizontal flip of a US image is still consistent with a 180-degree rotation of a US probe. A vertical flip, however, would result in dark shadow regions above a highly reflective structure, which is not in line with the physics of ultrasound. Rotating or translating the US image creates a dark gap between the transducer and the imaged tissue, which again does not align with ultrasound acquisitions, as the US probe has to touch the skin directly to create a US image.

Due to these differences between classical data augmentation techniques and the actual physics behind ultrasound, we propose novel augmentation techniques inspired by the physics of ultrasound that mimic true variations in US images.

6.3 Methodology

Our data augmentation approach is inspired by the physical model of wave propagation in US. Specifically, we model deformation caused by probe pressure, reverberation artifacts, and signal-to-noise ratio to create various augmented US images (Fig. 6.2). These augmentations require detecting regions of high echogenicity or attenuation, such as bones, and are currently designed for linear probe images.

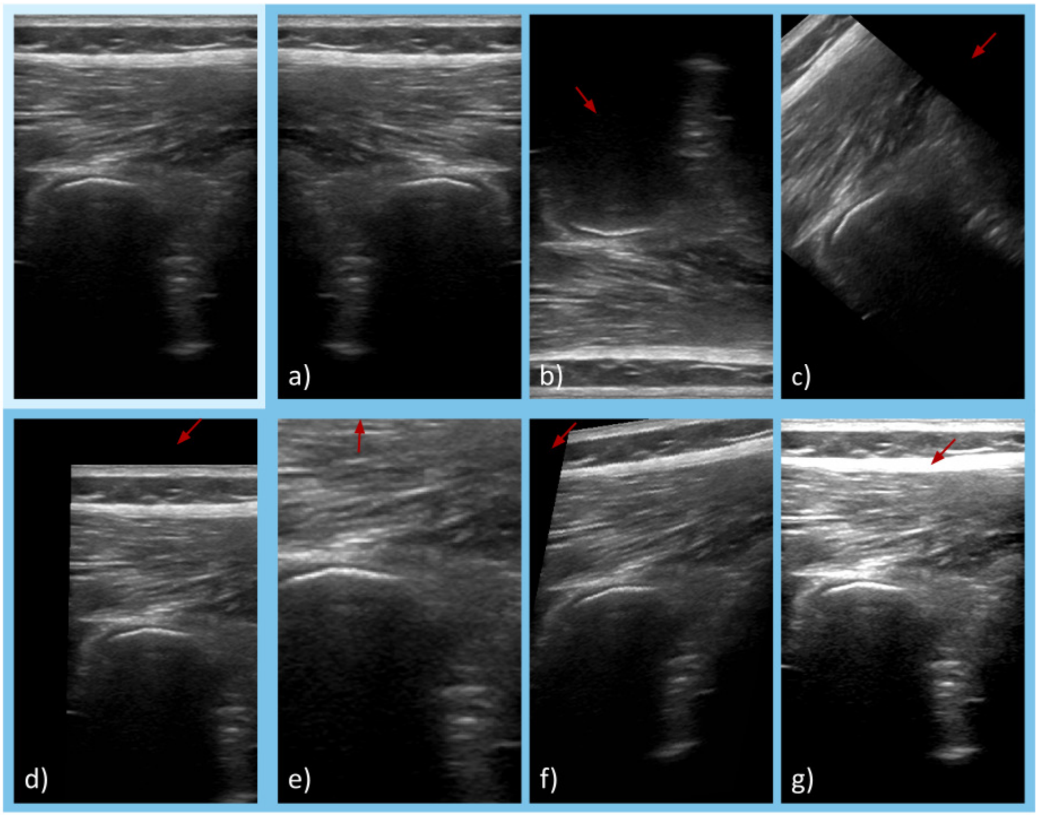


Fig. 6.1. Classical augmentations from the computer vision domain applied to a US image of a bone: a) horizontal flip, b) vertical flip, c) rotation, d) translation, e) scaling, f) shearing, and g) brightness modification. Adapted and reproduced with permission from Springer Nature [11].

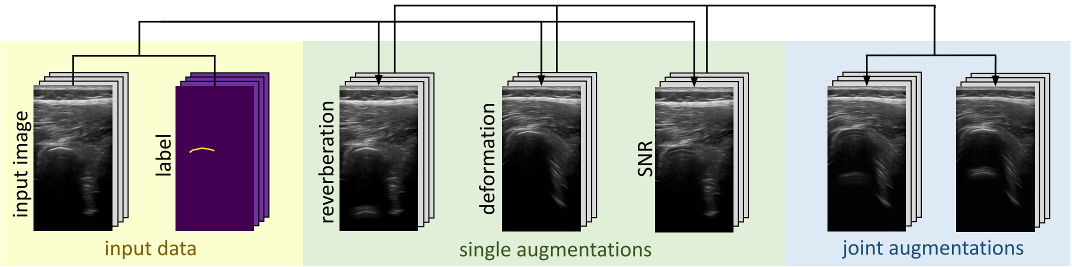


Fig. 6.2. The proposed data augmentation pipeline: from US images and corresponding bone masks (yellow), different augmentations are generated (green) and can be merged in the same image (blue). Reproduced with permission from Springer Nature [11].

6.3.1 Deformation

The proposed deformation augmentation models deformations arising from variations in the contact pressure of the probe on the imaged tissue (see Fig. 6.3). First, we derive a Displacement Field (DF) for the deformed medium based on the probe displacement d_{probe}. In this case, the DF is generated based on a simplified physical model of the interactions between soft tissues, bones, and the transducer. This model includes five assumptions for the previously mentioned interactions:

1. The soft tissues are isotropic and homogeneous.
2. The probe pressure is only applied in the axial direction.
3. Bones are modeled as static objects without deformation or transformations.
4. Only the tissue between the probe and the bone is exposed to compression.
5. Tissue deformation in the lateral direction is negligible.

Points 1-4 can be modeled by defining the deformation coefficients as

$$\epsilon_{yy} = \frac{F}{EA} \tag{6.1}$$

which can be rewritten to

$$F = d_{\text{probe}} \frac{EA}{y_{\text{probe}}} \tag{6.2}$$

Here, A defines the transducer area, F is the force applied by the transducer, E is Young's Modulus, and y_{probe} is the transducer position in the bone coordinate frame. Young's Modulus is a measure of the stiffness and deformation of a material, defined by the ratio between tensile stress and tensile strain. Based on these formulations, we can write the DF $\Sigma(x, y)$ as

$$\Sigma(x, y) = \int \epsilon_{yy} dy \bigg|_{\Sigma(x,0)=0} = -H(-y)\frac{F}{EA}y = -H(-y)\frac{d_{\text{probe}}}{y_{\text{probe}}}y \tag{6.3}$$

$H(-y)$ is the Heaviside step or unit step function. In this formulation, unrealistic gaps between the probe and skin can occur. To overcome this, we rewrite the DF in the probe coordinate system, which corresponds to applying an offset of d_{probe} to the initial equation such that

$$\Sigma(x, y) = -H(-y)\frac{d_{\text{probe}}}{y_{\text{probe}}}y + d_{\text{probe}} \tag{6.4}$$

Based on point 5, we can assume a constant DF in lateral tissues, equal to the probe displacement d_{probe}. Applying lateral gaussian smoothing ensures even transitions between regions with large and small displacements. Finally, the resulting DF is applied to the original B-mode image to create an augmented one. d_{probe} is sampled randomly to generate different augmentations.

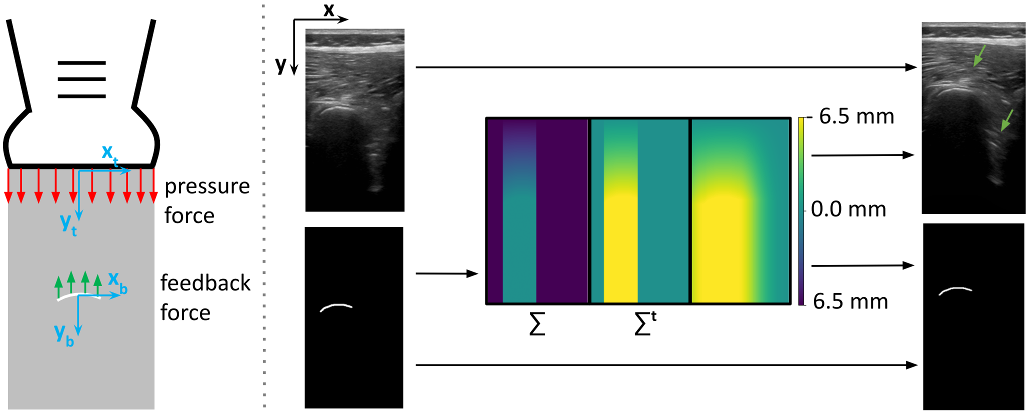


Fig. 6.3. Deformation Augmentation. Left: forces applied by the US probe (red arrows) and bones (green arrows), and the coordinate systems centered in the bone (x_b, y_b) and the probe centroids (x_t, y_t), respectively. Right: The original B-mode image and label serve as input. The label is used to compute the DF (Σ) induced by the probe pressure. To avoid an unrealistic gap above the skin layer, the DF is expressed in the probe coordinate system (Σ^t) and blurred. Σ^t is finally applied to both image and label. Adapted and reproduced with permission from Springer Nature [11].

6.3.2 Reverberation

Reverberation artifacts show a copy of highly reflective structures in deeper tissue areas. We model this behavior through a ray-based approximation and assume a constant speed of sound c in tissues. Then a highly reflective structure located at depth Δ will create an echo at

$$\Delta = \frac{ct_1}{2} \tag{6.5}$$

with t_1 being the time that a sound wave requires to leave and return to the transducer. In some cases, the echo intensity can be high enough to reflect the signal back into the tissue at the tissue-transducer interface. Then the signal is reflected again at the highly reflective structure, leading to an additional echo which is recorded at

$$\Delta_r = 2\Delta = t_1 c = \frac{t_2 c}{2} \tag{6.6}$$

with t_2 representing the duration between when the signal departs from the transducer and returns after a second reflection. Depending on the intensity of the echo, these multiple reflections can occur several times at multiples of depth Δ in the US image.

Figure 6.4 shows the proposed pipeline for this augmentation. First, we locate the bone centroid (x_c, y_c) in the label map. Then we extract the patch that contains the bone in the US image and move it by $2y_c$ in the axial direction. This shift creates an image that only contains the shifted bone further down in the US image. The shifted patch image is multiplied by a Gaussian filter in the reverberation region w_r and scaled by the reverberation intensity r_i. The multiplication weight for the original image is defined as $w_o = 1 - w_r$. r_i defines the intensity of the reverberation artifact and is sampled randomly. Lastly, the augmented image is generated through a sum of the original weighted image and the weighted shifted reverberation patch.

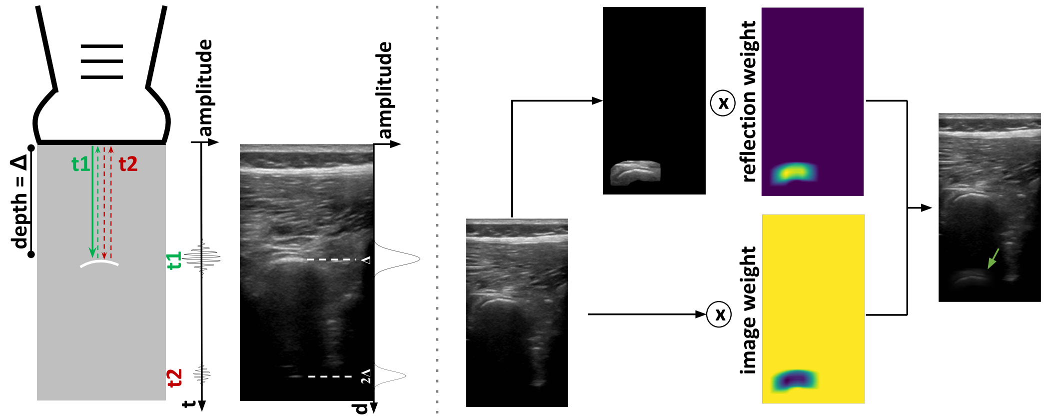


Fig. 6.4. Reverberation Augmentation. Left: The transmitted signal and the echos generated by a bone at depth Δ. The signals generated by the first echo (green arrow) and reverberation echo (red arrow) are shown in the time-amplitude plot. The first echo generates a bright area in the B-mode image at Δ, while the reverberation echo generates an artifact at 2Δ. Right: The patch containing the bone is extracted from the original US image and replicated at multiples of Δ. The replicated patch and the original image are weighted and summed up. Reproduced with permission from Springer Nature [11].

6.3.3 Signal-to-Noise Ratio

An image's signal-to-noise ratio (SNR) describes the relationship between the amount of desired signal and background noise. In the proposed augmentation (see Fig. 6.5), we propose to tune the SNR, thereby increasing or decreasing the visibility of bones compared to the background. First, we compute each B-mode's local energy (LE) map as proposed by Bridge et al. [119]. This LE defined as

$$LE(x) = f_e(x)^2 + f_{o1}(x)^2 + f_{o2}(x)^2 \tag{6.7}$$

provides information about the signal energy in each pixel in space. $f_e(x)$, $f_{o1}(x)$, and $f_{o2}(x)$ are all components of a monogenic signal which is extracted from US data. Then the US image is normalized by the corresponding LE map. Next, we create two tuned LE maps for bone and non-bone pixels by multiplying both regions with different scaling factors, i_b and i_{bg}, respectively. In this way, we can tune the relation between bone and background. If $i_b > i_{bg}$, bones are enhanced and vice versa. Lastly, the original US image is augmented through multiplication with the sum of both tuned LE maps. i_b and i_{bg} are sampled randomly.

6.4 Experiments and Results

We evaluated the proposed augmentation techniques against classical methods on a set of US images of the spine and by comparing network performances of segmentation and classification tasks.

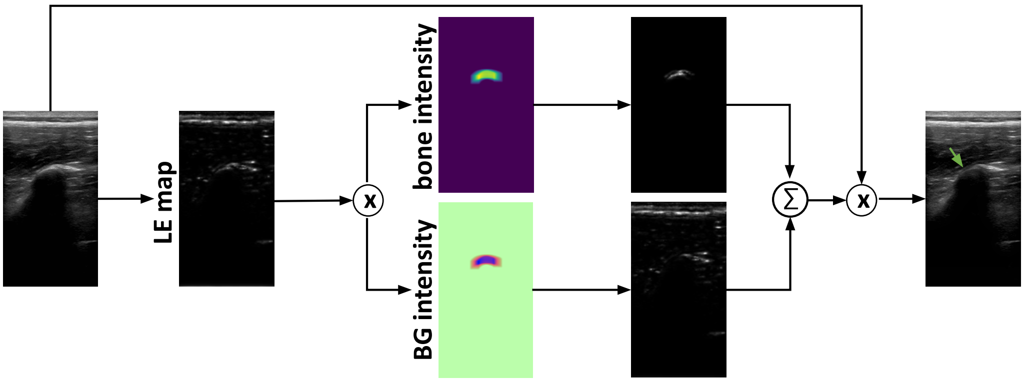


Fig. 6.5. SNR augmentation. The local energy (LE) map is computed from the US input image. The LE map is then multiplied with the label or its inverse and the respective intensity values to obtain tuned local energy maps for bone and background (BG) regions. The overall tuned local energy is computed by adding the bone and background parts. Lastly, the augmented image is obtained by multiplication of the input image with the tuned LE maps. Reproduced with permission from Springer Nature [11].

6.4.1 Dataset

To evaluate both deep learning models, we acquired US images from 24 healthy volunteers with a body mass index (BMI) ranging from 20 to 25. The scans were acquired with a Zonare z.one ultra Convertible Ultrasound System (Zonare Medical Systems Inc.) and an L8-3 linear probe. Furthermore, we set the following acquisition parameters: 7 cm image depth, 92% gain, 10 MHz frequency, and 13 MHz sampling rate. The acquisitions resulted in a total of 10.656 US images with a size of 272×503. This data is publicly available, split into the subsets used for segmentation and classification [1].

6.4.2 Segmentation

For the bone segmentation task, we applied a U-Net architecture [120] with a learning rate of 0.01, Adam optimizer, and Binary Cross Entropy loss function for training. We used 5.284 US frames containing bones and annotated the bones pixel-wise in each frame. We then split the dataset subject-wise in the following way:

- 3.972 frames from 17 subjects were used for training.
- 782 frames from 5 subjects were used for validation.
- 500 frames from 2 subjects were used for testing.

We trained the model with 5-fold cross-validation, and for evaluation, we used the Dice Score (DSC) and the Hausdorff distance (HDF). The DSC measures the overlap between the predicted and ground truth data, penalizing false negatives and false positives equally. The HDF measures how far apart two sets are from each other, in this case, predicted and ground truth segmentation.

---

[1] `https://github.com/mariatirindelli/UltrasoundAugmentation/tree/release/miccai2021`, visited on: 14.02.2023

Augmentation	**Parameter**	**Min**	**Max**
Classical	rotation in degree	-10	10
	translation (horizontal and vertical)	-0.2	0.2
	brightness	0.8	1.2
Proposed	deformation - d_{probe}	30	100
	reverb - r_i	0.50	0.9
	SNR - i_b, i_{bg}	0.70	1.40

Tab. 6.1. Value ranges for the classical and proposed augmentation parameters. For each augmentation, the parameters are uniformly sampled from the value range during training. The translation parameters describe the minimum and maximum fraction of translation based on the image width and height. Adapted and reproduced with permission from Springer Nature [11].

6.4.3 Classification

For the bone segmentation, we applied a DenseNet121 [121] with a learning rate of 0.01, Adam optimizer, and Binary Cross Entropy loss function for training. In this task, we used 5.692 US frames from 22 subjects. We then split this dataset into one class which contained at least one bone per frame, and the other that did not have bones in a frame. The data was split again subject-wise into training, validation, and testing, with the following amount of frames: 3.821 frames from 16 subjects, 1.037 from 4 subjects, and 834 frames from 2 subjects, respectively. We also trained the classification task with 5-fold cross-validation, and we evaluated the models' performances through average accuracy (ACC) and F1-score (F1). The ACC shows the average overall accuracy per class with the accuracy depicting how often the model predicted correctly over all predictions. The F1 score is the mean of precision and recall of a mode.

Table 6.1 shows value ranges for the applied augmentation techniques. For classical augmentations, we additionally used scaling and shearing. We then evaluated the network performance for each set of augmentations. We chose value ranges empirically from the given ranges. For both tasks, the augmentations were only applied during training and generated augmented images on the fly with 30% probability.

6.4.4 Results

The network performance for the segmentation and classification task with different augmentations are depicted in Table 6.2. The reader can see that the proposed augmentations, specifically deformation, and SNR, yield better results in the segmentation task. Deformation and SNR show the best dice score while using SNR augmentations results in the lowest Hausdorff distance. The proposed reverberation augmentation and classical augmentations resulted in the best network performance for classification. The average accuracy showed the best results for classical and reverberation augmentations, while using the latter produced the highest F1 score. Figure 6.6 shows examples of all proposed augmentation techniques.

Augmentations	Segmentation		Classification	
	DSC	HDF	ACC	F1
None	0.589 ± 0.07	20.72 ± 3.84	0.876 ± 0.06	0.770 ± 0.15
Classical	0.625 ± 0.03	17.76 ± 3.17	**0.883** ± **0.04**	0.780 ± 0.09
Reverb	0.604 ± 0.03	19.71 ± 2.20	**0.883** ± **0.03**	**0.802** ± **0.04**
Deformation	**0.626** ± **0.01**	19.06 ± 3.63	0.865 ± 0.04	0.759 ± 0.11
SNR	**0.626** ± **0.02**	**17.24** ± **1.83**	0.877 ± 0.06	0.764 ± 0.16
All proposed	0.600 ± 0.02	17.32 ± 2.97	0.834 ± 0.02	0.742 ± 0.04

Tab. 6.2. Performance comparisons for bone segmentation with UNet and bone classification with DenseNet. We report the average metrics over 5-fold cross validation along with their respective standard deviation (±). Adapted and reproduced with permission from Springer Nature [11].

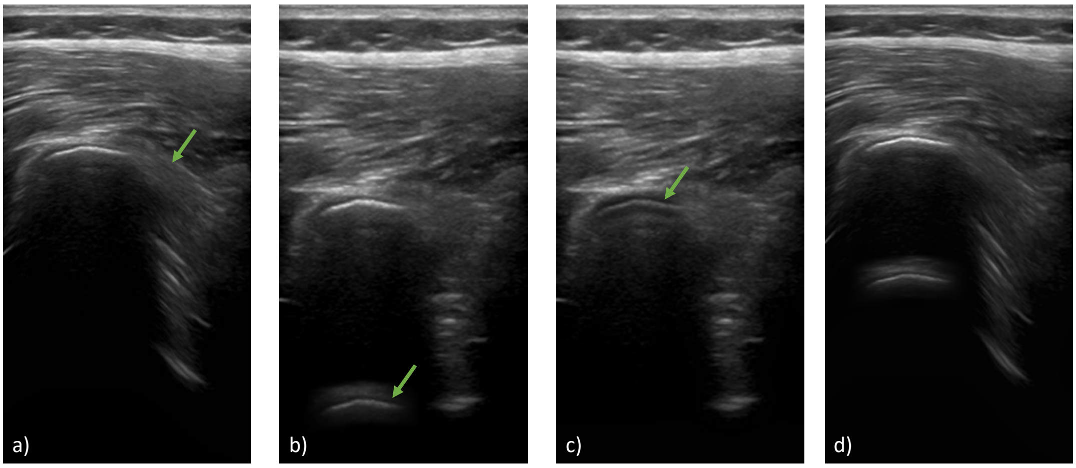


Fig. 6.6. Examples of the proposed US augmentations: a) deformation, b) reverberation, c) SNR, d) all proposed techniques combined.

6.5 Discussion

The network performance results lead to several learnings. First, we can see that any augmentation improves both the segmentation and the classification tasks. This result shows that increasing the variety of US images, in general, is beneficial. Furthermore, the proposed physics-inspired augmentations do not worsen the network performance. The low standard deviation among all folds emphasizes performance consistency when applying the proposed augmentations.

All evaluation metrics show a performance drop when combining all proposed augmentation techniques. One reason for this could be that the combined augmentations create images not present in the unseen test set. In the segmentation task, the dice score does not show significant differences when applying different augmentations. Therefore, the exact type of augmentation could be less important for improving the network performance. While the reverberation artifact does not improve segmentation, it improves classification. During reverberation, the visible bone is again visualized lower in the image. The segmentation could therefore fail at deciding which bright area is the bone. In contrast, the second bright structure can serve as an additional feature that distinguishes frames with bones from those without bones for classification.

Overall, the proposed augmentations do not significantly improve network performance for segmentation or classification in our dataset. However, applying physics-inspired augmentations produces more realistic synthetic US images. Because these augmentations are based on the physical behavior of ultrasound, we believe that the research community should use these augmentations and further developments in this area instead of classical augmentation techniques. In this way, this work can enable increased understanding and explainability of the network behavior in US imaging.

6.6 Future Work

The results highlight the potential of US-specific, physics-inspired augmentations. In the future, the research community should investigate further augmentation approaches. These approaches could be based on differences in the acquisition parameters, probe-tissue interactions, and scan acquisitions. Furthermore, the current pipeline requires bone labels which can be cumbersome to annotate. Future works could therefore use different automatic approaches to detect bones or propose functions without a specific anatomy detection. This approach would extend the augmentations to other US acquisitions, for example, those only containing soft tissue or muscles.

Now that we have seen how we can optimize the variability in US datasets, we will analyze and propose methods to optimize 3D US compoundings in the next chapter.

Robotic Ultrasound Compounding

7

Contents

In this chapter, we will look at one of the first steps for a complete robotic US framework: US compoundings. We will start with a short motivation and analyze the current state of the art. Then we will propose a new intelligent approach for US compoundings optimization, focusing on the compounding algorithm's input by mimicking a sonographer's approach to creating a mental anatomical model.

7.1 Motivation - The Mental Model

Ultrasound image compounding can increase interpretability [122] for downstream tasks such as segmentation and classification. Therefore, optimal 3D US compounding is crucial for improving medical imaging tasks. As mentioned in Sec. 2.4, a 3D compounding is an arrangement of two-dimensional frames in a 3D space, creating a 3D US volume. The translation into three-dimensional space is done through a compounding algorithm. Therefore, this approach can be optimized in two ways: by updating the compounding algorithm itself or by improving the input frames given to it. Several methods exist for the compounding task, and recent works focus on improving these algorithms. However, this thesis focuses on the latter option, improving the input for the compounding algorithm.

Analyzing the human body based on US information is a daily task performed by sonographers. To achieve this task, sonographers create a model of the anatomy of interest in their brain.

This process is done iteratively. First, the physician scans the area of interest to get a general overview, creating an initial mental model. However, some areas of this model are incomplete or imprecise. The sonographer, therefore, takes additional scans in these areas and, by default, updates the mental model. This process is repeated until the model is good enough for the sonographer to achieve the desired clinical task.

Such a process is logical and understandable for humans. However, it is performed by a human for whom the task is time-consuming, requires experience, and can lead to musculoskeletal pain in the long run. Therefore, in this chapter, we are asking ourselves how to use this knowledge to automate the process robotically. Which methods are necessary? And how can we tell whether our digital model is informative enough? The proposed approaches will answer these questions and give clues about possible future research directions to complete the task.

7.2 Related Work

As mentioned before, improving US compoundings can be performed in two ways: by improving the compounding algorithm and the input to the compounding algorithm. In this related work section, we will analyze the current state of both sides. First, we will look at current approaches to improve US compoundings in general. Afterward, we will look at recent robotic navigation approaches to improve US images. This second part includes works that focus on real-time probe adjustment and projects that preoperatively improve the US frame selection.

7.2.1 Ultrasound Compounding

US compounding can generally be seen as a combination of interpolation and averaging [123, 124] of B-mode frames. Averaging can be achieved through mean or weighted approaches. While this averaging can reduce noise, it simultaneously reduces detail in the final image. To counteract the loss of detail, weighted averaging can be applied. In this case, higher weights can, for example, be used in high signal-to-noise ratio (SNR) images. Researchers should choose the exact weighting method depending on the application and the imaging goals. Even though these averaging approaches are used frequently, both versions are prone to noise and are limited to static images. Therefore, further research is taking place to improve US compounding algorithms.

Some works, for example [125], suggest including phase data into the compounding algorithm. Phase data refers to the phase difference between different echoes that are reflected at varying parts of the tissue. This parameter depends on the distance to the reflecting surface and tissue properties. Phase data is computed based on the raw radio frequency (RF) data received by the US transducer. For the compounding task, phase information is taken into account through averaging or more complex information, such as the relative phase difference between images. This method can reduce noise and preserve edges and details in US compounding. Furthermore, it is less sensitive to tissue motion and can be applied to dynamic US imaging applications. However, this approach requires access to radio frequency data which most commercially available US systems do not provide. This technique is, therefore, interesting to analyze in a research setting but requires more time to be applied in clinical practice.

Ultrasound, by nature, is a view-dependent modality. Several works have therefore been proposed to consider view-dependency for US compounding. Computational sonography, introduced by Hennersberger [126] and Göbl et al. [127] uses elaborate view-dependant representations for 3D US compounding. At each location in 3D, the amount of reflected signal expected from each direction is encoded in a tensor, thereby enabling a direction-preserving compounding method. In the later work, a spherical representation extended the tensor representation. Hung et al. [128] recently introduced a compounding method that preserves helpful boundaries by combining Laplacian and Gaussian pyramids. While mean and median compounding algorithms are widely used, they reduce contrast information in the final compounding. This work aims to counteract this issue by applying Gaussian and Laplacian pyramids. Compoundings are computed at different frequency bands and scales. Then a contrast-maximizing task can be used at specific frequency scales. In this way, the algorithm can preserve contrast and direction in the compounding. The final compounding does not necessarily lead to the most accurate compounding but enhances the compounding by maximizing the contrast.

Lastly, zu Berge et al. [129] introduced an uncertainty- and orientation-based compounding method. In the first step, the acquired US frames are clustered based on orientation and proximity. Next, these clusters are fused into a compounding, taking uncertainty information into account. This approach can create more reliable compoundings because uncertain pixel values do not influence reliable pixel information.

7.2.2 Robotic Navigation and Frame Selection

Most robotic US scanning approaches focus on executing perpendicular sweeps [130, 131, 4] because, in general, a probe orientation perpendicular to the skin surface results in the best B-mode image. In most cases, an RGB-D camera records the patient, and the patient surface is computed as a point cloud to achieve this task. Then normal vectors from this point cloud are calculated. The robot can then execute a perpendicular trajectory to the surface by combining point cloud points with their normal vectors. Graumann et al. [132] extended a single sweep to multiple parallel sweeps. In addition to a point cloud, a region of interest (ROI) was also required. The robot then covered the area of interest with parallel sweeps. However, the robot still kept the probe perpendicular to the surface. Other works, such as Hennersperger et al. [65], showcased robotic navigation based on preoperative MRI data. The planning is done on MRI, and later the MRI is registered to a patient's point cloud. Works in this direction, however, require an additional imaging modality for trajectory planning.

Perpendicular scans generally lead to the best image quality. However, this case changes when obstacles are involved. Highly reflective structures, such as bones, reflect nearly all US signals. Therefore, structures behind the bone are not imaged, and the overall US image loses information. Thus, recent research also focuses on more intelligent scans to receive the most information from US sweeps and compoundings. Göbl et al. [67] introduced a US frame selection based on acoustic window information for liver scanning. On available CT data, the method computes a mean transmission estimate for target points in the liver from points on the surface by virtually modeling the transmission of US waves. Then the algorithm chose optimal transducer poses based on their estimated transmission to the target point. However, this work required preoperative imaging data from CT scans. Chatelain et al. [133] proposed an online robotic servoing based on confidence maps. The probe is kept orthogonal to the surface when no obstacles are detected. Once an obstacle is detected through confidence maps, the probe performs an in-plane rotation to avoid the obstacle. However, this real-time approach does not take the general structure of obstacles into account to avoid obstacles overall. Secondly, the image quality is improved by rotating away from the obstacles. Yet, it is not guaranteed that the volume around and beneath obstacles is scanned sufficiently.

7.3 Initial Digital Model Formation or Optimized Volume Coverage

With the initial model, we would like to get an overview of the region of interest without requiring information from a different imaging modality. The robot could execute a first scan for an initial model perpendicularly to the patient's surface. This approach can, however, be inefficient. Therefore, we propose a coverage approach that efficiently covers an unknown volume with as few transducer poses as possible.

To efficiently cover an unknown volume, this volume first has to be defined. In our case, the volume of interest is represented as a three-dimensional grid of voxels of arbitrary size (see Fig. 7.1, left). The top row of the voxel grid corresponds to the volume's surface; thus, the transducer is positioned on this surface to initiate scanning downwards. The transducer pose, denoted as t, is defined as a tuple comprising the tilt angle and the x-y coordinates of the voxel grid surface where the transducer is located, such that $t := (\theta, x, y)$. θ represents rotations of the transducer with respect to the surface normal. In theory, the transducer can rotate around all three axes. However, in the current work, we restrict θ to one degree of freedom, explicitly allowing for tilting in the out-of-plane direction. Furthermore, the current algorithm is set up discretely. Therefore, the tilt angles require upper and lower boundaries to calculate all possible rotations within the defined limits. Then the algorithm computes all possible transducer poses based on the defined grid and the available tilt angles.

We model ultrasound sweeps as linear interpolations between poses, specified by their grid locations and tilt angles. The transducer pose covers a voxel if the voxel intersects with the transducer's beam, which is modeled as an infinitesimally thin image plane. The coverage value of a voxel is 0 if it is not intersected. Otherwise, the coverage lies between 0 and 1, linearly proportional to the distance between the voxel center and the focal point of the transducer in that given pose. This focal point marks the point of the highest lateral resolution in a US image. The d function can then be interpreted as an inverse distance weight. Thus,

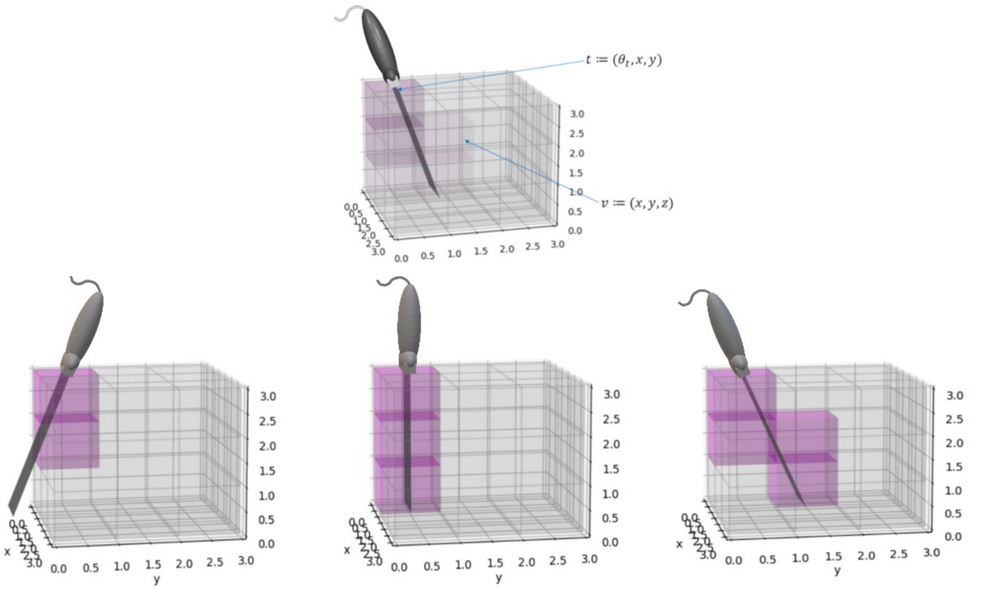


Fig. 7.1. Top (adapted from ©2022 IEEE [12]): basic grid structure used in the optimization. Each voxel v is specified by its x, y, and z coordinates within the grid. The transducer pose t is defined by an out-of-plane tilt angle, θ, and its location on the grid surface (x, y). Colored voxels represent voxels covered by the current pose. Bottom (adapted from [134]): different tilt angles in the same position lead to a two (left), three (middle) or four (right) covered voxels.

voxels closer to the focal point have a high d function value, and the further away a voxel is, the more its value decreases towards 0. Therefore, it holds that

$$d(t,v) = \begin{cases} f(t,v), & \text{if } v \text{ is intersected by } t \\ 0, & \text{otherwise} \end{cases} \tag{7.1}$$

with $t \in T$ being one of the considered poses and $v \in V$ being a voxel. $f(t,v) : (T,V) \to [0,1]$ is a function that linearly assigns a value between 0 and 1 depending on the distance from the voxel center of v to the focal point of the pose t.

Now that we have an initial setup, we can define an optimizer for our given task. The optimizer aims to determine the weight values, $w(t)$, that maximize the coverage value over the entire volume. Each weight estimates the information content and importance of a frame acquired with a specific pose in the volume reconstruction. For instance, if a frame acquired with a given pose $t(\theta, x, y)$ is imaging voxels that have already been covered, the optimizer will assign a relatively low weight to such pose as it is not adding any new information to the accumulated volume. The weights are then utilized to sort the transducer poses based on their contribution to the information content and overall volume coverage, thereby creating the weight vector w. It is worth noting that the optimizer minimizes given cost functions. Therefore, the coverage term is placed in a negative exponent in order to be maximized. The cost function of the volume coverage, $E_{VC}$is then given by:

$$E_{VC} = c_1 \cdot \exp(-\sum_{t,v} d(t,v) \cdot w(t)) + c_2 \cdot \sum_{t} w(t) + c_3 \cdot \sum_{v}(\sum_{t}(d(t,v) \cdot w(t) - \overline{cov_v})) \tag{7.2}$$

We define the average coverage over all voxels as the average of the coverage for each voxel, such that

$$\overline{cov_v} = \frac{\sum_v \sum_t (d(t,v) \cdot w(t))}{|V|} \tag{7.3}$$

with $v \in V$ being all voxels in the voxel grid and $t \in T$ being all possible transducer poses given the range of tilt angles and transducer locations. c_1 and c_2 represent separate weights for each cost function term. The last term in Eq. 7.2 represents the standard deviation of the coverage over all voxels. Once this term is minimized, the coverage is distributed more evenly throughout the volume.

The outcome of the cost function is a list of all possible transducer poses sorted by their contribution to the volume coverage cost function. To not execute all poses, the list has to be cut at some point. To determine this point, we introduce a cut-off function that discards remaining poses from the list once the average coverage over all voxels has reached a global threshold, indicating that the volume has been sufficiently covered on average. Additionally, this function ensures that all voxels reach a lower minimal coverage threshold, guaranteeing that each voxel is at least minimally covered.

After the optimization step and the cut-off, the ordered and reduced pose list is used to create a trajectory for the robotic arm to acquire the ultrasound sweeps. Images from this sweep are then used to compute the initial compounding.

7.4 Model Update

After constructing an initial compounding and an initial digital model, this model can contain unknown or low-quality areas that should be improved. While there are many reasons for low-quality areas in US compoundings, this thesis focuses on acoustic shadows. Arising from highly reflective structures, like bones, these shadows reduce the available information in a US image to nearly 0 below the bone. Therefore, we propose a method to locate these shadowed areas and improve the amount of information in these areas through rescanning.

7.4.1 Compounding Quality Assessment

We compute confidence maps on B-mode images to identify acoustic shadows as described in [135]. Confidence maps use a random walk method, where seeds are placed at the start of the transducer scanlines. For each pixel in the image, the confidence is calculated as the probability that a random walk starting from that pixel reaches a transducer element. Factors such as high horizontal edges or pixel depth can decrease this probability. Examples can be seen in Figure 7.2. By taking the confidence values of the pixels belonging to a voxel and thresholding the mean of these values, we can determine if that specific voxel is shadowed. Then by iterating over all voxels in the compounding, we can group the volume of interest into covered, good-quality voxels and shadowed low-quality voxels. In the following steps, the coverage of the low-quality voxels can then be improved.

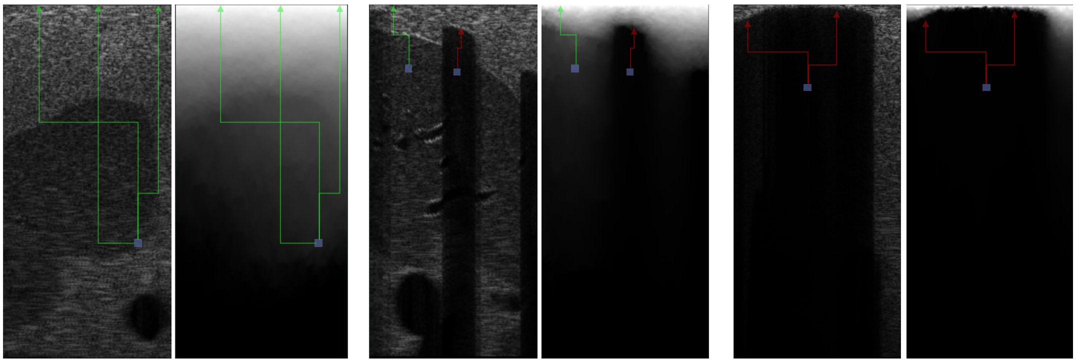


Fig. 7.2. Pairs of US images (left) and their correlating confidence maps (right). Green and red arrows show examples of random walks that reach or do not reach a transducer element from a specific pixel (blue), respectively.

7.4.2 Acoustic Shadow Reduction

In the next step, we reduce the amount of low-quality shadowed voxels arising from acoustic shadows. The so-called shadow reduction optimizer aims to identify poses with a high probability of covering shadowed voxels by using poses as different as possible from those known to lead to shadowed voxels. This approach assumes that distinctly different poses have a higher chance of not intersecting the same shadowed voxel in the same way as during the initial scan. The pose difference comprises two parts: the rotation angle differences and the distance between their (x, y) locations. In the current work, we are focusing on one rotational degree of freedom, the out-of-plane tilting.

In the first step, the tilt angle distances for each pose to shadowing poses are added to their d function value. In this way, poses with tilt angles very different from the tilt angles of known shadowing poses have a higher, therefore more preferable d function value. The cost function for the acoustic shadow reduction then consists of three parts: a modified version of the volume coverage cost function $E_{VC_{occ}}$, a similarity penalty term, and a voxel penalty term.

With the volume coverage term $E_{VC_{occ}}$, we aim to cover shadowed voxels from the first run as efficiently as possible, so the volume coverage terms are applied only to known shadowed voxels, leading to a preference for poses that efficiently cover these.

For the similarity penalty term, we penalize poses with (x, y) coordinates close to those of a known shadowing pose, while poses that originate far away from shadowing poses have a low penalty. This penalty is realized by defining a 2D Gaussian Mixture Distribution, where Gaussians are centered at the location that previously generated acoustic shadows. Then it holds that:

$$P_{sim} := \sum_{t} [w(t) \cdot g_{occ}(t)] \tag{7.4}$$

with

$$g_{occ}(t) = \left[\prod_{t_o \in O} \mathcal{N}(t_o, \sigma^2) \right] (t_x, t_y) \tag{7.5}$$

Here O refers to the set of known shadowing poses t_o, and t_x and t_y are the (x, y) location of the transducer pose t.

Lastly, the voxel penalty term penalizes poses depending on how well a specific voxel is already covered. Here poses are penalized more if the corresponding transducer beam intersects voxels with an already high coverage value. Then we can P_{vox} define as:

$$P_{vox} := \sum_t \left(w(t) \cdot \sum_{i,j} \left[\sum_{v \in V} d(t, v) \cdot \text{covered(v)} \right] \right) \tag{7.6}$$

with

$$\text{covered(v)} := \begin{cases} 1, & \text{if cov(v)} > \tau \\ 0, & \text{otherwise} \end{cases} \tag{7.7}$$

Therefore, the overall acoustic shadow reduction cost function is defined as

$$E_{ASR} = E_{VC_{occ}} + c_4 \cdot P_{sim} + c_5 \cdot P_{vox} \tag{7.8}$$

with c_4 and c_5 being weighting terms for the similarity and voxel penalty, respectively.

Similar to the volume coverage approach, this optimizer creates a list of possible next poses ranked based on priority. We, therefore, also introduce a cut-off function for this list. In this case, the cut-off value is reached in one of two cases: either a minimum of two new poses has covered all shadowed voxels, or the number of possible new poses for a specific voxel overstepped a global threshold. With the last condition, we ensure that unreachable voxels are ignored, and a cut-off will occur.

7.5 Experiments and Results

We performed a range of experiments to validate the proposed approach. These experiments are executed in simulation and on a real-life phantom with a robotic manipulator. We chose experiments in simulation to ensure a controllable environment and be able to analyze the effect of the proposed methods solely. The experiments in real-life serve as an initial evaluation of the approach in a more realistic setting. The following sections will describe the setup, specific experiments, and results. All sections are split into ’Simulation’ and ’Robotic’ parts to elaborate on both experimental setups separately.

7.5.1 Setup

Simulation

The majority of experiments were conducted in simulation to ensure a controllable environment. We generated the phantoms by placing high-density geometrical objects in a medium with a lower absorption coefficient (see Fig. 7.3). The generated robot trajectories were then fed to the ImFusion software for Ultrasound Simulations (ImFusion GmbH, Germany) on these synthetic phantoms to create US sweeps. This simulation is based on a ray-tracing model that generates ultrasound simulations from label maps, using a generative model to create speckle densities and intensities [136]. Descriptions of the different mimicked tissues can be found in Table 7.1. The high-density objects were simulated as bone, once placed in water and once in a soft tissue surrounding.

medium	**acoustic impedance**	**speed of sound**	**attenuation coefficient**
	(MRayl)	(m/s)	(MHz/cm)
bone	0.612	3600	7.9
water	0.149	1492	0.019
soft tissue	0.163	1540	0.54

Tab. 7.1. Tissue types and corresponding ultrasound parameters used in the US simulation.

Furthermore, we used the ImFusion software to compute confidence maps from the B-mode images of the hybrid ultrasound simulation. During simulations, the probe was modeled with a width of 32.5 mm, a central frequency of 2 MHz, an image depth of 10 cm, and a focal distance of 2.5 cm. The region of interest was defined as a bounding box within the label map. We set the voxel grid dimensions to $8 \times 10 \times 6$ voxels (LxDxH), and the list of considered tilt angles ranged from -35 to 35 degrees in 5-degree increments. We ran both the simulation and optimization process on a portable computer equipped with an Intel i7 processor and a GeForce GTX 1050 Mobile graphics card.

Robotic

In another experiment, we also evaluated the proposed approach on a robotic setup consisting of a robotic arm (LBR iiwa 14 R830, KUKA GmbH, Augsburg, Germany) and a Siemens Acuson NX3 ultrasound machine (Siemens Healthineers GmbH, Erlangen, Germany). A VF12-4 12MHz linear probe was attached to the end-effector of the robotic arm. The simulated sweeps were replaced by the sweeps obtained using the robot. The scanned phantom had a simple rectangular shape that cast a shadow in the surrounding water. Figure 7.7 shows this setup on the left.

7.5.2 Evaluation

We evaluated both approaches, the volume coverage for the initial model and the acoustic shadow reduction based on volume coverage, confidence, and intensity averages below

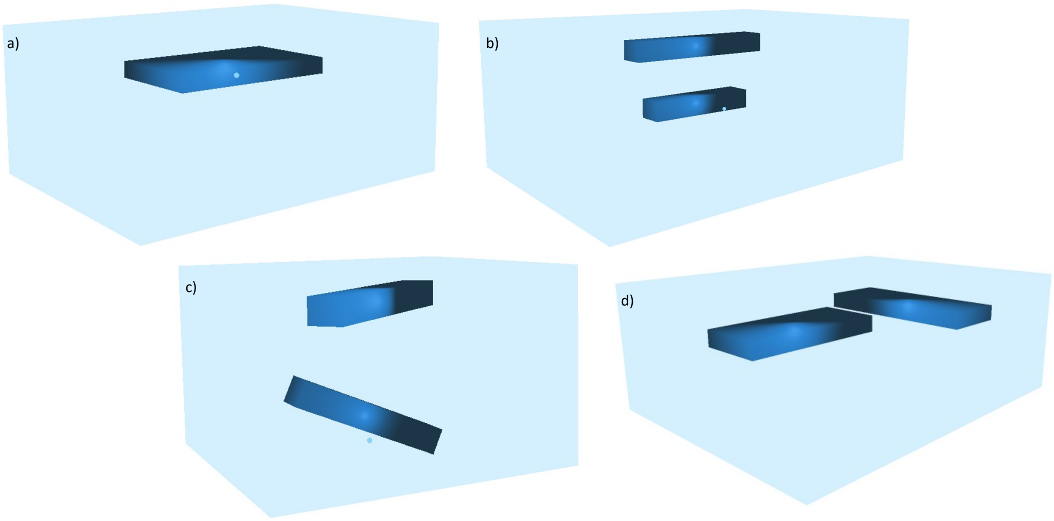


Fig. 7.3. Synthetic phantoms used for evaluation: a) square, b) parallel beams, c) non-parallel beams, d) non-connected plates. The light blue box shows the region of interest. For the US simulation, the dark blue parts were labeled as bone, while the light blue was once labeled as water and once as soft tissue (see Table 7.1). Adapted from ©2022 IEEE [12].

shadowing objects. The intensity values were obtained from two compoundings: one where the individual compoundings from both approaches were combined by averaging and the other where all sweeps from the respective strategies were combined before compounding them into a single volume. Our proposed method was compared against perpendicular, and random trajectories, with the latter generated using a quantum randomizer library for true randomness. The random scans generated different trajectories for each launch. Therefore, the values presented in the results are an average of ten runs for all configurations containing a random scan. The number of poses for the random scan was kept equal to the number of poses in the perpendicular scan, which is 80. We included these two comparisons because perpendicular scans generally produce the best image quality, and random scans have shown high performance in similar tasks as described in [137]. The acoustic shadow reduction was also applied to both perpendicular and random scans to analyze its impact on differently acquired initial compoundings. Alongside each iteration, the number of required poses is also reported. For the robotic acquisitions, we also state the scanning time.

7.5.3 Results - Volume Coverage

For the volume coverage, the weighting terms $c1$, $c2$, and $c3$ were set to 1.0. Therefore, all terms equally impact the optimization.

Simulation

Figure 7.4 shows examples of shadowed voxels after the initial scan (top row) and the acoustic shadow reduction (bottom row). We can see that the proposed volume coverage already leads to the smallest amount of shadowed voxels in the first compounding. Table 7.3 shows quanti-

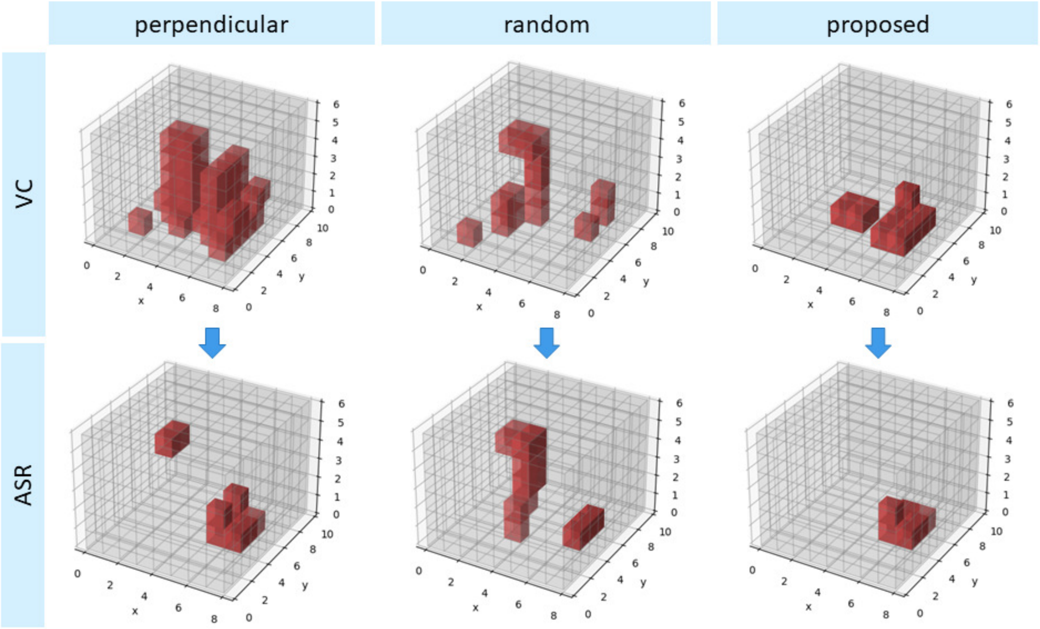


Fig. 7.4. Example of shadowed voxels after the initial volume coverage (top row) and the acoustic shadow reduction (bottom row) for perpendicular, random and the proposed sweep method. It can be seen that the proposed initial scan method results in the lowest amount of shadowed voxels after the initial compounding. The acoustic shadow reduction then reduces the number of shadowed voxels for all initial volumes. Adapted from ©2022 IEEE [12]

tative results for the volume coverage and the acoustic shadow reduction approach. Looking at the volume coverage alone, we can see that the average coverage and average confidence are highest with our proposed volume coverage approach compared to perpendicular and random scans, both for soft-tissue and water cases. However, the number of required poses is more than five times higher. Regarding intensities, perpendicular and random scans each show favorable results in two phantoms. Figures 7.5 and Fig. 7.6 also shows qualitative results of the volume coverage model (top row), highlighting that the proposed volume coverage approach also visually results in the best covered compounding for the two presented phantom cases.

Robotic

As shown in Table 7.2, the proposed volume coverage results in increased average coverage and average intensity of the initial scan. However, the number of poses is much higher than in the perpendicular or random approach. Figure 7.7 shows qualitative results of the different acquisition approaches. It can be seen that the proposed volume coverage (last column, first two rows) shows fewer dark or unscanned areas in the compounding than perpendicular or random scans.

7.5.4 Results - Acoustic Shadow Reduction

To decide on the values for $c4$ and $c5$ we empirically analyzed different values and combinations, resulting in setting $c4$ to 2.0 and $c5$ to 500.0.

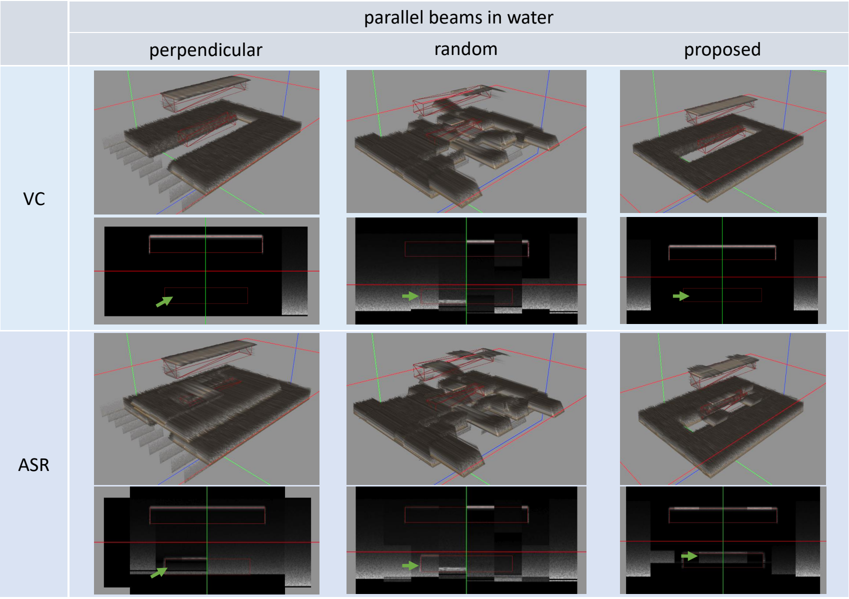


Fig. 7.5. US compoundings for the parallel beams phantom in water. The columns display results for perpendicular, random, and the proposed initial scan. The rows show volume coverage (VC) and acoustic shadow reduction (ASR) results. Both a 3D compounding and a 2D US slice are presented for each case. The green arrows indicate regions of successful acoustic shadow reduction. We show the final results for the perpendicular and the proposed initial scan because these methods are deterministic. For the random case, one possible outcome is shown. Adapted from ©2022 IEEE [12].

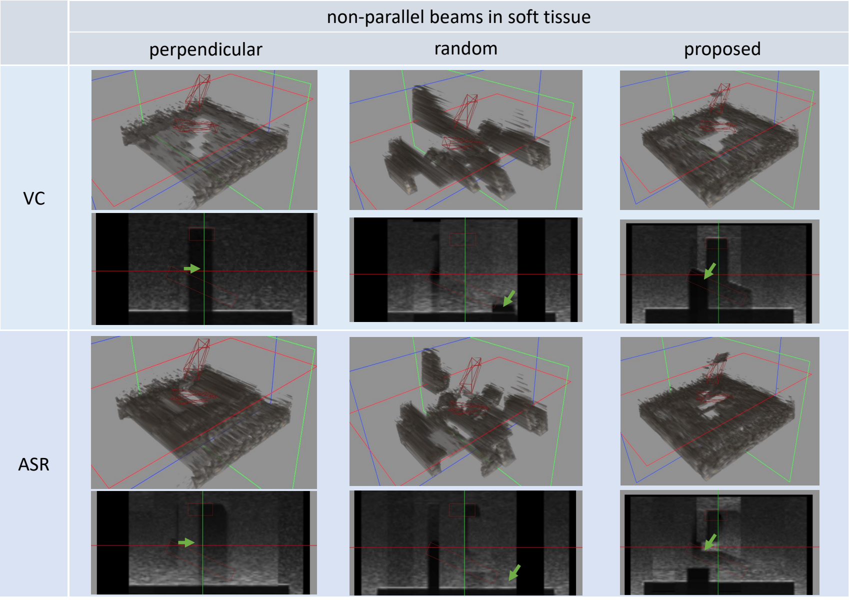


Fig. 7.6. US compoundings for the non-parallel beams in soft tissue. The columns display results for perpendicular, random, and the proposed initial scan. The rows show volume coverage (VC) and acoustic shadow reduction (ASR) results. Both a 3D compounding and a 2D US slice are presented for each case. The green arrows indicate regions of successful acoustic shadow reduction. We show the final results for the perpendicular and the proposed initial scan because these methods are deterministic. For the random case, one possible outcome is shown. Adapted from ©2022 IEEE [12].

method	avg. coverage	avg. confidence	number of poses	scanning time (s)
perpend.	0.425	0.052	40	**65**
random	0.213	0.024	**37**	97
VC	**0.4375**	**0.064**	176	88

Tab. 7.2. Experiment results for the robotic experiments on real data. ©2022 IEEE [12].

Simulation

The second row of Figure 7.4 shows the remaining shadowed voxels after the acoustic shadow reduction is applied to all initial compoundings. It can be seen that after the acoustic shadow reduction, the amount of shadowed voxels reduces compared to all initial volumes. The most significant reduction can be observed between the initial volume from perpendicular scans and the corresponding result after acoustic shadow reduction. In contrast, our proposed initial scan method already results in a small number of shadowed voxels, so the difference after the acoustic shadow reduction is also smaller. Table 7.3 shows the results for the acoustic shadow reduction in simulated soft tissue and water. It can be seen that the acoustic shadow reduction consistently improves the overall volume coverage and average confidence in all of the tested phantoms and for all acquisition protocols. Additionally, for three out of four phantoms, the acoustic shadow reduction step leads to higher intensity values in the regions located beneath the shadows.

The results in Table 7.3 also demonstrate that, on average, the combination of volume coverage optimization with the acoustic shadow reduction optimization yields the best coverage and volume confidence results. However, it requires a higher number of poses compared to other acquisition protocols. Both random and perpendicular scans utilize a similar number of poses. Nonetheless, perpendicular scans show better volume coverage and volume confidence results than random scans, as depicted in Figures 7.5 and Fig. 7.6. Using the shadow reduction optimizer with both perpendicular and random scans improves acquisition performance while maintaining fewer poses than the volume coverage method. On average, the number of iterations required for the volume coverage and shadow reduction optimizers is 26.422 and 86.472, respectively.

Robotic

We can extend Table 7.2 by performing our proposed acoustic shadow reduction approach on all initial compoundings (see Table 7.4). This table shows that the acoustic shadow reduction method increases the average coverage compared to all initial scans. The average confidence is increased after perpendicular and random scans but is kept equal after the volume coverage approach. The number of additional poses is the same for initial perpendicular and random scans, while the initial coverage compounding requires significantly more poses.

phantom	method	soft tissue					water		
		avg. coverage	avg. confidence	avg. intensity	combined intensity	avg. number of poses	avg. coverage	avg. confidence	avg. number of poses
Square	perpend.	0.921	0.442	21.520	21.323	80	0.750	0.352	80
	perpend. + ASR	**0.958**	**0.460**	**22.187**	**28.021**	138	**0.942**	**0.441**	191
	random	0.598	0.287	**31.267**	**29.655**	80	0.585	0.255	80
	random + ASR	**0.754**	**0.362**	25.007	29.014	162	**0.741**	**0.332**	167.7
	VC	0.983	0.472	**9.963**	9.712	448	0.954	0.454	448
	VC + ASR	**0.992**	**0.476**	5.280	**10.695**	468	**0.954**	**0.454**	505
Plates	perpend.	0.894	0.429	10.033	9.020	80	0.833	0.400	80
	perpend. + ASR	**0.975**	**0.468**	**14.831**	**20.750**	106	**0.900**	**0.432**	174
	random	0.441	0.212	**27.914**	**26.775**	80	0.428	0.188	80
	random + ASR	**0.514**	**0.247**	16.465	23.841	113	**0.525**	**0.232**	129.5
	VC	0.973	0.467	**10.794**	10.518	448	0.883	0.424	448
	VC + ASR	**0.981**	**0.471**	6.067	**11.460**	474	**0.900**	**0.432**	609
Beams, parallel	perpend.	0.915	0.439	4.825	3.822	80	0.850	0.404	80
	perpend. + ASR	**0.975**	**0.468**	**8.757**	**14.664**	116	**0.917**	**0.436**	166
	random	0.339	0.163	**28.930**	27.735	80	0.579	0.236	80
	random + ASR	**0.499**	**0.240**	22.354	**28.722**	138.4	**0.740**	**0.300**	146.6
	VC	0.992	0.476	**8.268**	7.266	448	0.950	0.456	448
	VC + ASR	0.992	0.476	3.688	**8.941**	460	**0.965**	**0.460**	499
Beams, non-parallel	perpend.	0.890	0.427	**15.517**	14.466	80	0.840	0.390	80
	perpend. + ASR	**0.904**	**0.434**	13.021	**20.924**	211	**0.917**	**0.436**	166
	random	0.567	0.272	**15.106**	13.701	80	0.628	0.260	80
	random + ASR	**0.757**	**0.363**	12.760	**16.058**	195.7	**0.773**	**0.325**	155.3
	VC	0.910	0.437	**12.935**	11.827	448	0.950	0.440	448
	VC + ASR	0.910	0.437	10.849	**16.484**	577	**0.970**	**0.450**	490

Tab. 7.3. Evaluation results on soft tissue and water simulation for volume coverage (VC) and acoustic shadow reduction (ASR) on all four phantoms. Random scans are averaged over 10 runs. ©2022 IEEE [12]

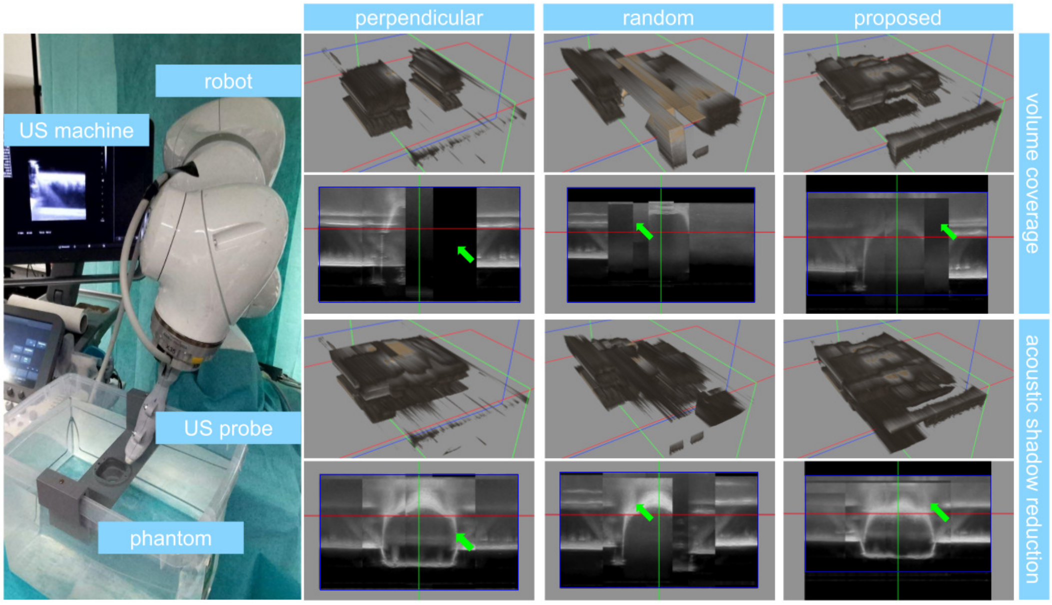


Fig. 7.7. Left: robotic setup consisting of a robotic manipulator, a US machine, a US probe and a geometric phantom. Right: qualitative results of the volume coverage and acoustic shadow reduction, comparing perpendicular scans, random scans and the proposed approach. ©2022 IEEE [12].

7.6 Discussion

Based on the results, we can see that our proposed method, modeling a sonographer's approach, is working and improving US compoundings. The initial model, meaning the initial compounding, shows better coverage results than perpendicular or random scans. Therefore, we achieve a better general overview of the volume of interest. In the second step, we rescan the volume of interest to reduce acoustic shadows and thereby increase the amount of information inside the compounding. The results show that this approach is also successful in improving compoundings.

Using the acoustic shadow reduction optimization technique could, in some instances, result in lower intensity values than only using the initial coverage step. This phenomenon can be attributed to the newly selected poses by the acoustic shadow reduction optimizer, which effectively accesses previously occluded voxels, but simultaneously occludes previously visible voxels through the shadow. Averaging the overall intensity can thereby decrease the intensity in initial brighter cases. Furthermore, the acoustic shadow reduction optimization improves volume voxel visibility. However, it may result in darker regions in the compounded volume below the shadowing surface. Both darker region phenomena can be mitigated by discarding shadowed voxels during the volume compounding process or by implementing weighting techniques as previously proposed in [129].

We can also see that the method performance partially depends on the number of poses. However, this argument should also consider the number of possible rotations. A perpendicular scan already covers all poses in the (x, y) space. Therefore, adding additional poses with the same orientation would not provide any additional information to the compounding.

method	avg. coverage	avg. confidence	number of poses	scanning time (s)
perpend.	0.425	0.052	40	65
perpend. + ASR	**0.433**	**0.058**	71	137
random	0.213	0.024	37	97
random + ASR	**0.279**	**0.036**	71	186
VC	0.4375	**0.064**	176	88
VC + ASR	**0.438**	**0.064**	211	175

Tab. 7.4. Experiment results for the robotic experiments on real data. ©2022 IEEE [12].

A possible way to reduce the number of poses would be to include this parameter in the optimization, thereby optimizing for low-quality voxels and the number of poses.

In our simulation and experimental setup, we considered flat acquisition surfaces, which might not be accurate for real-world clinical scenarios. However, it is important to note that our optimization pipeline is independent of the surface geometry. Additional visual sensors, such as RGB-D cameras, can create a mesh of the scanned surface, which can also be curved. Then normal vectors can be extracted from the mesh, and rotation angles can be based on these normals.

7.7 Future Work

Our sonographer-like approach for compounding optimization shows promising results. Nevertheless, future steps can be taken to improve the optimization. Initial future steps include the usage of curved surfaces and the introduction of additional rotation angles. Furthermore, the number of additional poses should be considered in the optimization process, such that this number is reduced compared to the current optimization. The updated optimization scheme can first be tested in a realistic medical simulation environment and afterward transferred to real-life cases. In both simulation and in the real world, it would be promising to analyze the performance of the proposed approach for liver US scanning. This specific task is physically and mentally challenging for physicians due to ribs covering large amounts of the liver. Exchanging the physician's approach with our robotic method while approaching the task like a physician could therefore reduce musculoskeletal pain and fatigue for physicians and improve diagnostics for patients.

Now that we have seen how to improve US compoundings, in the next part, we will see how to use such US compoundings for a specific clinical task.

Part V

Robotic Ultrasound Diagnostics

So which clinical procedure can be improved through intelligent, robotic systems to make these systems applicable?

In this part, we will answer this question by looking at thyroid diagnostics, specifically thyroid volumetry. This volumetry estimation is a routine clinical task executed manually by physicians with the help of US imaging. However, as already mentioned in the Introduction, manual 2D US has many drawbacks, which we want to counteract with an intelligent robotic system. We will first give a small medical introduction to the thyroid and its diseases. Then we will look at current works in thyroid segmentation and robotic thyroid scanning. Afterward, we will propose a novel robotic and automatic thyroid volumetry framework and compare this method to the current clinical gold standard.

Robotic Ultrasound of the Thyroid

8

Contents

8.1 Motivation

To make robotic US applicable to a clinical task, we first have to analyze the medical and clinical background. In the motivation, we will first give an overview of the thyroid, its functionalities, and possible diseases. Next, we will look into current diagnostics and treatments for thyroid diseases which require ultrasound imaging. We will see that ultrasound is a frequently used and highly required imaging form for thyroid diagnostics; therefore, introducing robotics could benefit this clinical area.

8.1.1 The Thyroid and its Diseases

The thyroid is a butterfly-shaped organ located in the neck, right below the larynx, and partially wrapped around the trachea. It consists of two lobes, one on either side of the trachea, and the isthmus, which connects both lobes. The thyroid produces hormones with different functionalities, including regulation of the metabolic rate, control of the heart, digestive and muscle functions, bone maintenance, and brain development. To achieve this task, the thyroid needs a sufficient amount of iodine. If this supply is ensured, follicular cells in the thyroid absorb the iodine and, in return, create thyroid hormones. Furthermore, the thyroid is connected to the brain network to ensure regulated hormone production.

Most diseases that are related to the thyroid can be mainly divided into two groups: thyroid hormone production or nodules. Nodules can either be benign or malignant. Diseases regarding hormone production can again be divided into two groups: hyperthyroidism and hypothyroidism. In the latter case, the thyroid is underactive and does not produce enough hormones. This deficiency becomes visible as tiredness, weight gain, and muscle weakness, among others. During hyperthyroidism, the exact opposite effects take place. The thyroid is overactive and produces too many hormones, leading to symptoms such as nervousness, sweating, and increased and irregular heart rates. This disease is common in up to 2.5% of the population in iodine-sufficient countries and in up to 2.9% in iodine-deficient countries. Next, we will see how some of these diseases, specifically hyperthyroidism, can be diagnosed and treated.

8.1.2 Thyroid Diagnostics

Ultrasound is a well-established imaging modality in thyroid diagnostics. It is used to examine the thyroid and nodule tissue through parameters like echogenicity. Furthermore, US is used to measure the thyroid and nodule volume. This latter parameter is an important clinical factor. It is used to observe the thyroid over a longer period to assess whether a therapy is working. Furthermore, it is used in radioiodine therapy (RIT), a treatment for hyperthyroidism. In this procedure, a small dose of radioactive iodine, typically ^{131}I, is administered to the patient via oral intake, with most of the radioactive material concentrated in the thyroid gland. The dosimetry in this procedure is based on the Marinelli formula (see Eq. 8.1), which requires the thyroid mass (m in gramm), the target dose of ^{131}I to the thyroid (D in Gy), its 24-hour uptake (IU_{24h}, unit-less) and its effective half-life time in the thyroid gland (T_{eff} in hours) [138].

$$A = \frac{25 \cdot m \cdot D}{IU_{24h} \cdot T_{eff}} \tag{8.1}$$

in which 25 is a unit conversion coefficient and A represents the administered activity in GBq. The thyroid volume is then used to infer the thyroid mass m by assuming constant density.

The current clinical standard combines 2D US with an ellipsoid formula to compute the thyroid volume. To achieve this task, the physician first scans both thyroid lobes separately for an overview. Then the physician moves the probe to the location where he or she can see the biggest area of the thyroid, both in transverse and sagittal view (see Fig. 8.1). In both views, the physician then measures the longest distance of the thyroid lobe in each visible direction, and these measurements then serve as an input to the ellipsoid formula. The procedure is then repeated on the other lobe to get two lobe volumes and one thyroid volume by adding both sides together. Research, however, has shown that 2D US is difficult to analyze and highly user-dependant [4]. Moreover, this user dependency can be extended to the volumetry task due to its dependence on 2D US. Therefore, we propose a framework that combines robotic

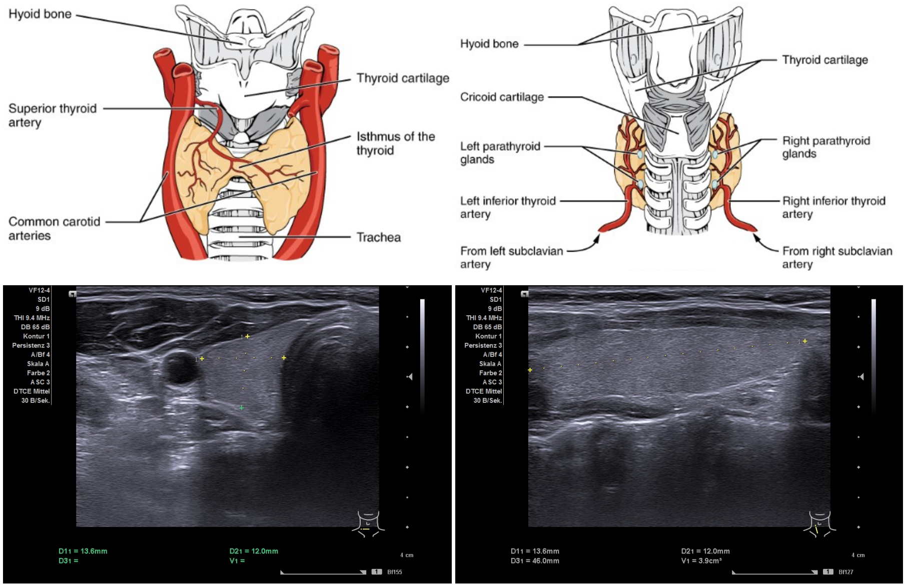


Fig. 8.1. Left: depiction of the thyroid and its surrounding anatomical structures from anterior (top) and posterior view (bottom) [1]. Right: both 2D US views for measuring the thyroid volume. The dotted yellow lines show the measurements. The probe orientation on the neck is shown on the bottom right of each US image.

acquisitions and automatic segmentations to increase accuracy and reduce user dependency of thyroid volumetry.

8.2 Related Work

In this section, we will focus on two different research areas that have not been combined before our proposed approach. First, we will look at the current state of automatic thyroid segmentation in US images. These projects introduced methods for segmenting the thyroid in 2D and 3D US data. Next, we will look at robotic US scanning of the thyroid. Here we will see navigation and controller design approaches for robotic US scanning, which are evaluated on thyroid scanning.

8.2.1 Thyroid Segmentation

Thyroid segmentation in ultrasound is a vast research topic. In 2020, Chen et al. [139] published a review article on this topic. This work divided segmentation methods into contour- and shape-based, region-based, and machine- and deep-learning methods. Furthermore, the segmentation task can be applied to 2D and 3D US. Segmentation methods also differ based on the amount of user interaction. Some semi-automatic methods still require user input. These methods can, for example, increase the efficiency of ground truth segmentations. At the same time, many deep learning segmentation methods now function fully automatically.

---

[1] `https://commons.wikimedia.org/wiki/File:1811_The_Thyroid_Gland.jpg`, visited on: 14.02.2023

In previous work [140], we combined a deep neural network with tracked 3D US data for thyroid volumetry and compared it to the current clinical gold standard. The tracked 3D data was acquired through electromagnetic tracking and converted into a 3D compounding. Then a QuickNAT [141] was used for the thyroid segmentation. Results on volunteer data showed that combining tracked 3D US with an automatic segmentation decreases intra- and interobserver variability. Furthermore, we saw that the automatic approach showed the most considerable improvement for the least experienced physician.

Lastly, publicly available data plays a crucial role in any deep learning approach, but publically available US data is rare. For the previous project, we gathered tracked US scans of 32 volunteers. To foster research on thyroid US and increase comparability between deep learning approaches, we made the dataset publically available [2]. It consists of tracked US scans of 32 volunteers, ground truth segmentations, MRI scans of the neck, and MRI segmentations of the thyroid. The project is continuously being updated so that it currently also contains segmentations of the jugular vein and the carotid artery, both in US and MRI. Internally, we also acquired a dataset of tracked 3D US for 186 patients. This dataset is currently being annotated and will be made publically available once the annotation is finished.

8.2.2 Robotic Thyroid Scanning

Previous studies have explored the use of robotics in US thyroid gland imaging. Kim et al. [142] developed a control algorithm for robotic US scanning utilizing position and force feedback. The system controlled the applied force by analyzing the root mean square error (RMSE) between consecutive images. A support vector machine was then trained to optimize the image quality through force adaption based on the RMSE input. Their work was evaluated on a thyroid phantom by a clinician ranking the final 2D US images. Huang et al. [68, 69] developed automatic scanning systems using both a 3D translating device and a robotic arm. In both cases, two force sensors were directly attached to the front of the US probe and used for force control during scanning. For path planning, the team first extracted a point cloud of the patient's surface, defined a region of interest, and then computed contours in this region. Finally, the robotic device scanned this ROI with the required amount of parallel sweeps. After scanning, a 3D reconstruction of the US volume was computed. The systems were evaluated, on a thyroid phantom, amongst others, by analyzing the image quality of the 3D reconstruction qualitatively. Kaminski et al. [70] conducted a feasibility study of robot-assisted ultrasound scanning for thyroid disease assessment. The robotic system employed force feedback control and a pre-registered anatomical model to plan the scanning trajectory. Acquired 2D B-mode images were then compounded into a 3D volume. The scan repeatability was evaluated by calculating the cross-correlation of images obtained from multiple runs. Kojcev et al. [4] compared measurements of thyroid length in 3D obtained through robotic acquisition to manual 2D expert measurements. The robotic acquisition was achieved through force feedback and a manual definition of the region of interest on the acquired patient point cloud. Their evaluation showed that measurements on robotically acquired 3D volumes are more repeatable compared to the manual 2D approach.

---

[2] `https://www.cs.cit.tum.de/camp/publications/segthy-dataset/`, visited on: 14.02.2023

Despite the demonstrated effectiveness of robotic US scans for the thyroid, these studies all involve significant human involvement in trajectory planning. Moreover, these approaches do not utilize the acquired data to adjust or refine the trajectory for 3D reconstruction of the thyroid. Finally, only Kojcev et al. use the acquired data for a clinically relevant task.

8.3 Methodology

In this section, we will describe the detailed methodology of the proposed idea. First, we will give a general framework overview. Then we will introduce our robotic navigation, consisting of the basic movement and movement corrections, both based on the US image and live segmentation. Next, we will describe the network structure for automatic thyroid segmentation. Last, we will present the volumetry estimation based on the segmentation results.

8.3.1 Framework Overview

The overall framework consisting of two steps can be seen in Figure 8.2. Before the US scan acquisition, the US probe is manually placed on the neck by the operator so that the thyroid is visible in the US image. This step is the only part requiring human interaction. The robot then moves towards one end of the lobe based on online segmentation results obtained using a QuickNAT network [141]. Once the segmentation does not detect thyroid tissue, the first end of the thyroid lobe is reached. Then the recording of ultrasound images and probe positions starts as the robot moves toward the opposite end of the thyroid lobe. The system adjusts its planned path in real-time based on live ultrasound images and the corresponding segmentation to ensure consistent image quality and reliable acquisition of the entire thyroid volume. The second part, the automatic volume estimation, begins after the US scan has been performed on both lobes. Both lobes are combined into one volume, and the trained network is applied to the acquired scan. The resulting segmentation is then used for a volumetry analysis. Furthermore, the acquired 3D US scan can be used for additional analysis, such as nodule detection.

8.3.2 Robotic Navigation

The proposed robotic navigation consists of a baseline navigation and motion corrections, which will be described here in detail. The baseline movement enables the scanning of the thyroid from the starting point to one end and back to the other end of the thyroid lobe. Simultaneously, the motion corrections enhance the US image quality in each step by removing side shadows and centering the thyroid lobe in the image.

Basic Movement

Internally, the robotic manipulator is controlled by a built-in real-time impedance control (KUKA Sunrise.OS 1.15.0.9). Through this control, the robot executes constant force onto the

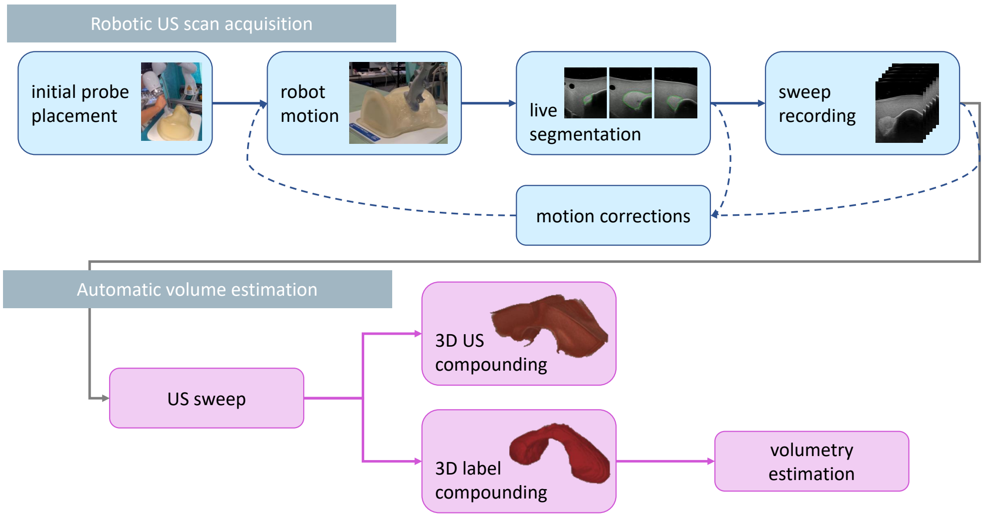


Fig. 8.2. Framework for robotic thyroid scanning with automatic thyroid volumetry consisting of an automatic robotic US scan acquisition and an automatic volume estimation. Adapted from ©2022 IEEE [13].

phantom in the z-direction while moving freely in the x-direction. Furthermore, this control allows retracting the probe from the phantom easily.

On top of this impedance control, we add image-based navigation and motion adjustments. After manual placement of the probe on the neck, the initial scan trajectory is planned along a line perpendicular to the probe orientation (see Fig 8.3, left, pink arrow). The robot then advances in 0.5 cm increments along the x-axis. After each step, the real-time segmentation is analyzed to assess the presence of thyroid tissue in the current US image. If all segmented images in the last second contain thyroid tissue, the system considers the thyroid to be present. Once thyroid tissue is not segmented continuously, the robot reverses its direction, and recording the US poses and B-mode images starts. Upon not detecting thyroid tissue continuously for a second time, the recording and the scan finished, and the probe retracted from the neck.

Motion Corrections

Apart from the initial scan based on the online segmentation of thyroid tissue, we introduced motion correction algorithms in the US image, namely shadow prevention and lobe positioning (see Fig. 8.3,right). Both adjustment steps are executed in real-time iteratively after each movement step if needed.

In cases when the probe only partially touches the surface, shadows can arise on one or both sides of the US image. We counteract these shadows by rotating around the movement axis (x-axis). First, the presence of shadows is analyzed through an intensity-based method. The left and right margins of the image are each divided into equal segments. We detect a shadow if a specified percentage of pixel intensities in all segments on one side is below a designated threshold value (p_b). The probe is then rotated step-wise by an angle $\alpha_{\text{corr,step}}$ in the direction of the shadow side on the image. This process is repeated until the percentage of bright

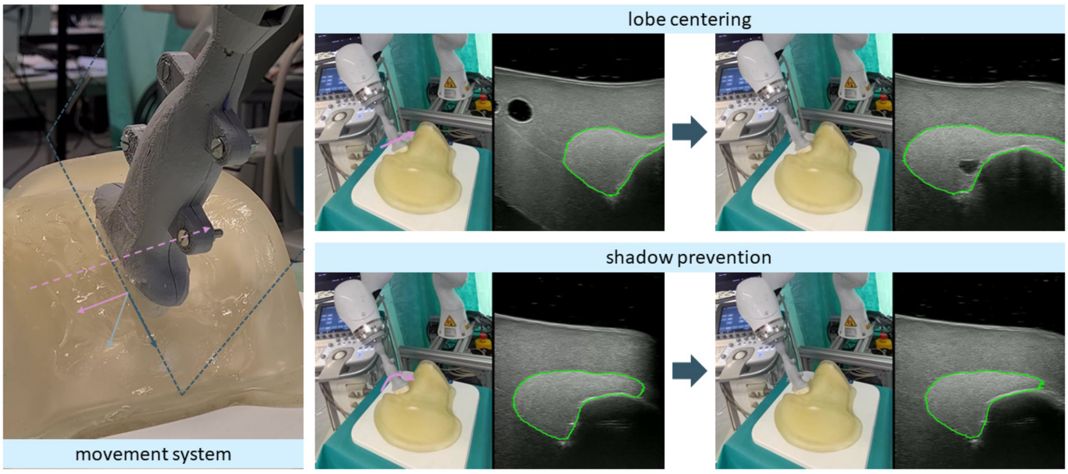


Fig. 8.3. The movement definitions and motion corrections. Left: the probe moves in x-direction (pink) and can rotate and translate in the y-z plane (dark blue dotted). Right: robot and US image examples before (left) and after (right) the motion corrections. ©2022 IEEE [13]

pixels in each border segment is bigger than p_b, resulting in a total correction angle of α_{corr}. If shadows are detected on both sides, the system does not perform any rotational adjustment but pauses and sends a warning to the user.

To achieve accurate volumetry estimations, we want to cover the thyroid as much as possible, thereby covering the biggest amount of thyroid tissue possible in each US frame. For this task, we introduce a sideways correction (y-z plane) to achieve maximum coverage per frame. This lobe positioning is done in the following way. The segmentation can be seen as a binary label map of thyroid tissue and background. Then we can determine whether the thyroid label exists in the left or right side of the label map or on both sides. If only one border of the image contains a thyroid label, the probe is translated sideways in the y-z plane. The direction and amount of incremental adjustments are defined so that the outer edge of the thyroid label is positioned close to the respective vertical image border. The total correction in the y-direction is then defined as y_{corr}. If both edges of the thyroid contain the thyroid label, we follow a slightly different approach. In this case, we compute the center of mass of the segmented thyroid in the image and translate the probe so that the center of mass is at the center of the image. This approach ensures that wider thyroids can also be covered as efficiently as possible. In both cases, we allow for a defined margin between the thyroid label and the edge of the image sides to prevent excessive motion corrections.

Combining both motion corrections with the initial scan, we can define the desired pose using the following formulas. First, we define the correction rotation matrix R_{corr} based on the rotation adjustment angle, such that

$$\mathbf{R}_{\text{corr}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\left(\alpha_{\text{corr}}\right) & -\sin\left(\alpha_{\text{corr}}\right) \\ 0 & \sin\left(\alpha_{\text{corr}}\right) & \cos\left(\alpha_{\text{corr}}\right) \end{bmatrix} \tag{8.2}$$

Then the target probe rotation R_{probe} and translation t_{probe} can be defined by

$$\mathbf{R}_{\text{probe}} = \mathbf{R}_{\text{corr}} \cdot \mathbf{R}_{\text{init}} \tag{8.3}$$

$$\mathrm{t}_{\text{probe}} = \mathrm{t}_{\text{init}} + \mathbf{R}_{\text{init}}[s \cdot n, y_{\text{corr}}, 0]^{\mathrm{T}} \tag{8.4}$$

with R_{init} and t_{init} being the initial probe orientation and position after the initial placement. n defines the current step position based on the initial placement, and s defines the length of one movement increment, currently set to 0.5 cm.

8.3.3 Thyroid Segmentation

To automatically segment the thyroid in the US scan, we rely on the QuickNAT proposed by Roy et al. [141]. This network is similar to U-net [120] but can be applied to smaller datasets. The architecture, depicted in Figure 8.4, consists of dense blocks of four encoders and four decoders with a bottleneck layer in between the encoder and decoder part. Each dense block includes a batch normalization layer and a Rectifier Linear Unit followed by three convolutional layers. Two of these layers have a 5×5 kernel size, and one has a kernel size of 1×1. The bottleneck block is a $5x5$ convolutional layer with a batch normalization layer. The skip connections between the corresponding encoder and decoder blocks allow for direct transfer of encoder feature information to the decoder. Furthermore, training is improved through gradient flow to deeper layers. 2×2 max-pooling blocks after each encoder reduce the spatial dimensions of the feature maps, while un-pooling layers are located after each decoder block. The final layer presents a classifier block with softmax activation. This layer maps the input to a feature map with the number of channels related to the number of segmentation classes. In our case, we work with two channels: thyroid and background. The softmax then translates the channel information into probability maps for each class.

In this work, the segmentation was performed real-time on each B-mode image. These segmentations are then used for navigation and for volumetry. The segmentation can be performed on 2D or 3D US, meaning on images or in a compounding. On our acquired thyroid dataset [140], we compared a 2D U-Net [120] on US images with a 3D U-Net version applied on the corresponding US compounding. We noticed that the 2D version resulted in slightly better absolute results. However, the 3D version produced smoother predictions and had a smaller overhead.

8.3.4 Volumetry Estimation

After the US scans of both lobes have been acquired, the data is post-processed. In the first step, the 2D frames are fused into a compounding. All recorded 2D US images have to be correlated to their respective probe poses through timestamps for this task. Since probe poses and US images are not sampled regularly, the probe poses are interpolated to match the US image time stamps through linear interpolation in the case of probe translation and spherical interpolation in the case of probe rotation [10]. Once we have correlated image and pose pairs, these are used as input for a compounding algorithm, resulting in a 3D US volume.

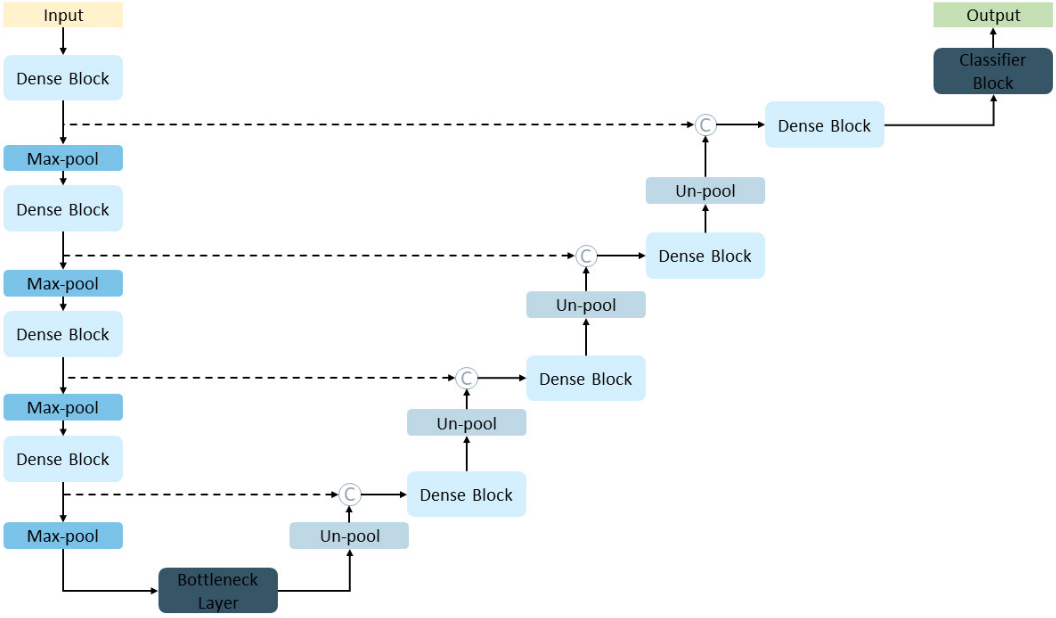


Fig. 8.4. Illustration of the QuickNAT architecture, based on [141].

The before-mentioned steps are also applied to the 2D thyroid labels, thereby creating a 3D thyroid segmentation.

This procedure is applied to both sweeps. Afterward, both volumes are merged by a spatially correct union of both sweeps and segmentations. The combined thyroid segmentation is then used to compute the thyroid volume by adding each thyroid label voxel. Therefore, the volume $V_{thyroid}$ can be defined as

$$V_{\text{thyroid}} = d_1 \cdot d_2 \cdot d_3 \cdot \sum_{i=1}^{l} \sum_{j=1}^{w} \sum_{k=1}^{h} M_{\text{thyroid}}(i, j, k) \tag{8.5}$$

with d_1, d_2, and d_3 as voxel dimensions being returned by the compounding algorithm. $M_{thyroid}$ is the 3D segmentation mask with

$$M_{\text{thyroid}}(i, j, k) = \begin{cases} 1 & v_{i,j,k} \text{ contains thyroid} \\ 0 & \text{otherwise.} \end{cases} \tag{8.6}$$

where $v_{i,j,k}$ represents a voxel in the segmentation mask and i, j, and k are the dimensions of this mask.

8.4 Experiments and Results

8.4.1 Setup

The system setup is similar to the one described in Chapter 5, consisting of a robotic manipulator (KUKA LBR iiwa, KUKA AG, Augsburg, Germany) and a Siemens Acuson NX3 US machine. The robot is equipped with a VF12-4 12MHz linear probe at the end-effector and controlled through ROS and iiwa stack [20][3]. B-mode images are captured through a frame grabber with 30 fps via a USB interface. Furthermore, we developed a plugin in ImFusion Suite Version 2.13.9 (ImFusion GmbH, Munich, Germany) to control and visualize the proposed framework.

8.4.2 Experiments

The experiments were performed on a thyroid phantom (thyroid ultrasound training phantom, model 074, CIRS, USA). Next, some parameters for the automatic robotic approach had to be defined. To enable shadow prevention, both sides of the images are divided into eight segments each. The margins are defined as 5% of the total image width. The intensity threshold p_b was set to 70, and a shadow is detected if 90% of the pixel intensities in each segment are smaller than p_b. Rotation steps to counteract a detected shadow are defined by the step size $\alpha_{\text{corr,step}} = 5°$ and limited to $\alpha_{\text{corr,max}} = 30°$. To reposition the lobe inside the US image, margins of 4% of the total image width are analyzed on both sides. If the lobe is not centered, the probe will move step-wise until the outer edge of the lobe segmentation is located at 6% of the total image width away from the image border with applying a maximum offset of $y_{\text{corr,max}} = 8cm$.

To assess the clinical value of the proposed method, we evaluated the automatic approach against the current 2D US gold standard. For this study, we recruited 12 volunteers (5 medical experts and 7 non-experts). The manually segmented thyroid on a CT of the phantom (voxel size $0.625 \times 0.625 \times 0.8$ mm^3) served as ground truth. Its volume amounted to 30.01 ml.

8.4.3 Results

Here, we will qualitatively and quantitatively analyze the effect of the motion correction algorithms. Next, we will look at the volumetry results from our proposed approach and the conventional manual 2D US method. We will analyze accuracy, repeatability, and dependence on experts for both cases.

Movement Corrections

We performed an ablation study in order to evaluate the effectiveness of the proposed motion correction methods. A total of 74 initial probe positions were recorded from 37 US scans of two lobes each. The phantom was then scanned in four different configurations with the same starting position: no motion correction, only shadow prevention, only centering, and both, resulting in 296 scans and 148 complete thyroid volume estimations. The fused and compounded segmentation masks were used to calculate the thyroid volumes for all runs.

---

[3] `https://github.com/IFL-CAMP/iiwa_stack`, visited: 14.02.2023

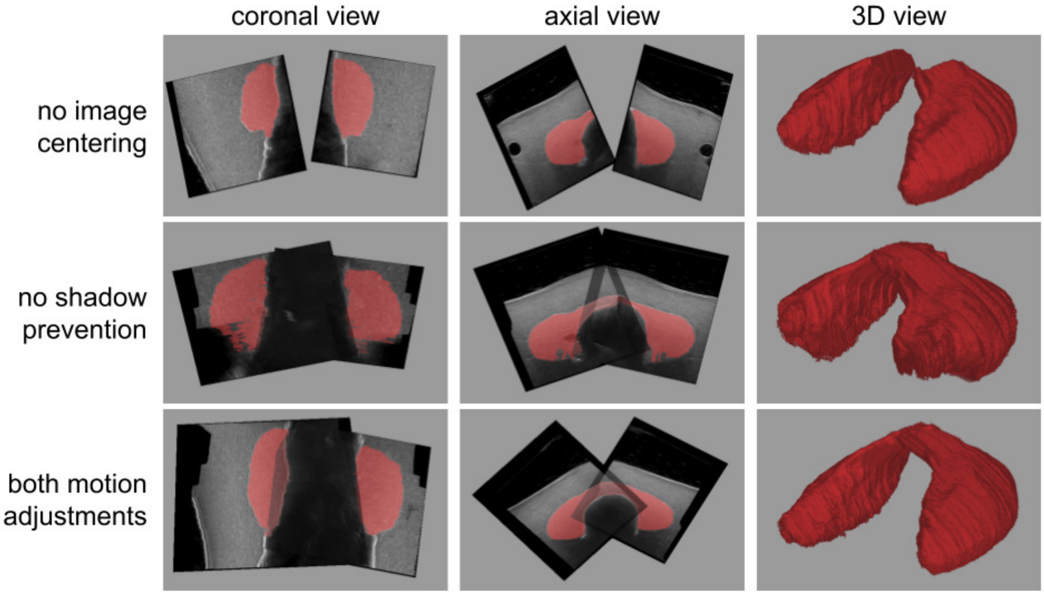


Fig. 8.5. 3D segmentations and 2D coronal and axial views of the thyroid with only shadow prevention (top), only lobe centering (middle) and both adjustments combined (bottom). ©2022 IEEE [13].

Quantitative results can be seen in Table 8.1. It shows that the volume estimation improved from having no motion correction to having shadow prevention and between applying the shadow prevention and both corrections jointly. However, only applying the lobe positioning resulted in worse volumetry estimations. We also show qualitative results in Figure 8.5. These images highlight that both motion adjustments combined create optimal 3D segmentations. When we do not apply lobe centering, the isthmus is not scanned entirely, resulting in missing parts in the segmentation. On the other hand, omitting shadow prevention leads to fuzzy segmentations at the end of both lobes.

type of motion correction	average volume [ml]	±std [ml]
none	29.30	3.63
only shadow prevention	**30.97**	1.69
only centering	29.05	4.60
shadow prevention & centering	31.90	**1.00**

Tab. 8.1. Volumetry estimations based on different movement corrections. ©2022 IEEE [13]

Comparison to Conventional 2D US Volumetry

Participants measured the thyroid volume with the automatic and the conventional approach. In the latter one, participants were instructed to measure the longest line in each of the three dimensions in the thyroid lobe in their selected 2D US images. An example can be seen in Figure 8.1 on the right. One lobe volume was then calculated using an ellipsoidal formula such that

$$V = c \cdot m_1 \cdot m_2 \cdot m_3 \tag{8.7}$$

with $c = 0.48$ as a correction factor and m_1, m_2, and m_3 as the measured lengths in all directions. This process was repeated for the second lobe, and the total thyroid volume was computed by adding both lobe volumes.

For the robotic approach, participants were asked to place the US probe on the phantom such that the thyroid lobe was visible and relatively centered in the image. The robot then scanned the lobe according to the previously described process. Each participant carried out both the conventional and the proposed methods three times.

Results of this comparison are presented in Table 8.2. They indicate a significant reduction in mean error from 20.85 ± 16.10 ml to 8.23 ± 3.10 ml in volume estimations between the conventional 2D US and the proposed robotic method in the expert group. In both groups, the average volume is more closely aligned with the measured ground truth, and the standard deviation is smaller, indicating a closer agreement between the volumetric estimates. Additionally, the difference in average volume estimations between experts and non-experts is marginal with the proposed robotic method.

method	**experts**	**non-experts**
conventional (ml)	23.75±4.83	18.56±4.47
robotic (ml)	**32.48±0.93**	**31.48±1.29**
ground truth (ml)	30.01	30.01
mean measurement error (conventional) (%)	20.85±16.10	38.16±14.89
mean measurement error (robotic) (%)	8.23±3.10	5.76±4.30
mean error improvement with the robotic method (%)	61	85

Tab. 8.2. Comparison between conventional 2D volumetry and our proposed method with the CT segmentation as ground truth. Volume is given as mean±std (ml). ©2022 IEEE [13].

8.5 Discussion

In the discussion, we will debate the influence of motion corrections and the possible clinical impact of the proposed method.

8.5.1 Movement Corrections

The results show that both movement corrections combined outperform not applying any movement correction. However, only applying the translation adjustment leads to worse performance. This behavior could be attributed to the introduction of shadows. During the centering of the thyroid lobe, the robot generally performs a translational movement toward the center of the neck. This motion can then result in a lack of contact of the probe onto the surface if not enough pressure is applied, thereby creating shadows on one side of the US image. More sophisticated movement directions can be developed to counteract this behavior. Furthermore, a real-time force adjustment could improve performance in these cases.

8.5.2 Volumetry

In contrast to the current state of the art in robotic US scanning of the thyroid, we combined robotic US scanning with a deep neural network to improve the volumetry task. Kojcev et al. [4] compared thyroid volumetry in conventional 2D US to 2D projections on robotically acquired 3D compoundings. While the results are promising, the physicians' manual measurement still influenced the volume estimation. Therefore, a factor for increased user dependency remains.

Both manual and robotic volumetry results were compared to a CT ground truth segmentation. To analyze the consistency of this segmentation, we analyzed the variability in this segmentation. Medical imaging experts created eight segmentations in the CT of the thyroid phantom. The mean and standard deviations of these segmentations amounted to $29.54 \pm 0.59 mm^3$. This standard deviation is significantly lower than the one in the US acquisitions. Therefore, variability in the CT segmentations is neglectable regarding the comparison to US results, and it is reasonable to use one CT segmentation as ground truth.

Apart from robotic tracking, 3D US compoundings can be obtained in different ways, such as optical or electromagnetic tracking. In these approaches, the US probe is tracked using optical or electromagnetic tracker. However, the scanning process is still carried out manually. Furthermore, accuracy measurements show that the robotic acquisitions have a better precision [143] compared to optical [144] or electromagnetic tracking [145]. In a previous study [140], we combined this electromagnetic tracking with the presented QuickNAT architecture for thyroid volumetry in volunteers and achieved promising results. However, the manual component in this setup can still lead to variability due to operator differences, human fatigue, and other factors. Combining this knowledge with the comparison in this work, we can reason about the order of different volumetry techniques. The current conventional 2D US approach is the least accurate and creates the most considerable user dependency. This result can be attributed to an inaccurate ellipsoid approximation, manual probe placement, and length measurements. Next, we can name the combination of manually tracked 3D US with an automatic deep learning segmentation. This approach already significantly improves the volumetry estimations and decreases user dependency. The increase in accuracy could be attributed to thyroid segmentation. An accurate thyroid segmentation is closer to its actual shape than an ellipsoid approximation, resulting in a more accurate volume prediction. Lastly, the proposed robotic approach should be named. While the segmentation accuracy can only be slightly improved, the user dependency is reduced even more compared to manual 3D tracking. Furthermore, as seen in the results, the robotic approach is also accurate when used by non-experts. This shows that in contrast to the conventional and manually tracked 3D approaches, the robotic approach does not require expert knowledge and can therefore be applied more widely.

The current study was evaluated on a thyroid phantom. Therefore, the current network is overfitted to phantom US images. However, in [140], we have shown that the network can successfully segment the thyroid in human subjects. We, therefore, believe that the segmentation-based navigation approach will also perform well on human subjects. Physicians mentioned that the thyroid phantom shows less heterogeneous information than a human neck structure. Therefore, the physicians could not rely on anatomical landmarks when measuring the thyroid volume of the phantom. This might be another reason for the significant discrepancy between manual and robotic volumetry results. We, therefore, believe that the gap between both approaches could be smaller when evaluating both on human subjects.

8.6 Approach Automation

A fully autonomous system does not acquire input from the operator. Therefore, instead of placing the US probe manually, the system has to approach the neck automatically. One way to achieve this would be by recording images and depth information from the neck through a camera attached to the robot and creating a surface mesh of the neck. However, this approach creates an entire neck surface mesh even though only two target points are required, one for each lobe. A leaner approach, therefore, consists of key point detection in the images and combining it with depth information for a specific point.

For this approach, we used OpenPose [146][4] to detect human body key points. OpenPose can detect 135 key points on the human, body, hand, face, and foot in real-time and on single images. In our setting, the US probe should approach the neck so that the US image in the target position shows thyroid tissue. We, therefore, first extract the nose, neck, and shoulder key points. Then we apply a geometric approximation between these key points to compute the x and y coordinates of the target point such that:

$$\begin{aligned} p_{\text{target,x}} &= p_{\text{neck,x}} - |p_{\text{shoulder,x}} - p_{\text{neck,x}}| \cdot o_v \\ p_{\text{target,y}} &= p_{\text{neck,y}} - |p_{\text{nose,y}} - p_{\text{neck,y}}| \cdot o_h \end{aligned} \tag{8.8}$$

where o_v and o_h are offset values in x and y axes, respectively. $p_{\text{target,x}}$ and $p_{\text{target,y}}$ are the target coordinates in the x- and y-axis, $p_{\text{shoulder,x}}$, $p_{\text{neck,x}}$, $p_{\text{nose,y}}$ and $p_{\text{neck,y}}$ are respective x and y coordinates of the extracted shoulder, neck, and nose key points. The z coordinate of the target point is provided through depth information from the RGB-D camera.

Initial results showed that this approach automatically places the probe on the neck with thyroid tissue visible in the US image. Combining this approach with the previously mentioned robot navigation can increase the autonomy of the proposed robotic thyroid US scan.

8.7 Future Work

During thyroid acquisitions on patients, we realized that neck anatomies show high variability. Therefore, the network has to be retrained to account for variable thyroid US images. Furthermore, these differences might require additional path corrections for robotic navigation.

---

[4] `https://github.com/CMU-Perceptual-Computing-Lab/openpose`

[5] `https://viso.ai/deep-learning/openpose/`

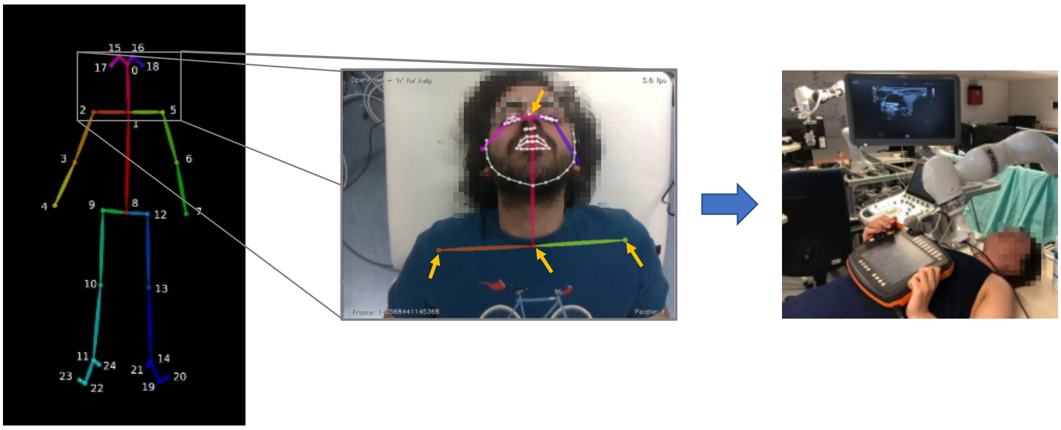


Fig. 8.6. OpenPose for the automatic approach. Left: human body key points that OpenPose can detect [5]; middle: we extract nose, neck, and both shoulder key points (yellow arrows); right: the robot successfully moves to the target point on the neck.

Some thyroids, for example, continue below the clavicle bone. In this case, imaging above the bone will not produce valuable images. In contrast, the probe has to be tilted below the bone so that the obstructed thyroid tissue can be imaged. Other patients exhibit goiters that change the surface to a convex shape. The navigation should be adapted to these differences in curvature to enable good imaging and ensure patient safety. More complex shadow prevention algorithms and force control can be included to improve the scan quality further. This study was conducted on a phantom. Next, studies should evaluate the system on human volunteers and, if successful, be tested on patients.

The shown automatic approach to the neck shows promising results. However, it can be improved further. The camera requires a distance of around 40 cm to detect key points and gather depth information. In shorter distances, this task does not perform accurately anymore. Therefore, we assume that the patient does not move in the current setup. Future software and hardware improvements could be analyzed to account for patient movement. Moreover, in some cases, the US probe touched the patient's chin during the approach. Head detection could therefore be included in the path planning to avoid collisions between the probe and the patient. Lastly, it could be possible to switch to deep learning-based landmark detection algorithms to compute the target points on the neck.

Now that we introduced intelligent and applicable robotic US systems, we will analyze their acceptance and factors that can influence this acceptance in the next part of the thesis.

Part VI

Patient Acceptance

How can we ensure that patients accept intelligent and applicable robotic systems?

In this part, we will try to answer this question by analyzing the effect of interactions and communication on acceptance. First, we will report relevant work in acceptance and trust in medical robotics. Then we will propose an interactive and communicative framework for robotic imaging. We will evaluate this framework in robotic US scanning and assess different evaluation metrics for acceptance.

9 Patient Acceptance in Robotic Ultrasound

Contents

9.1 Motivation

Acceptance is crucial in most human interactions, both in everyday situations and in clinical settings. When patients visit a physician, they accept their expertise and ability to diagnose and treat their condition. This acceptance also extends to the clinical procedures that the physician performs. In these cases, statistical outcomes of clinical procedures are known and therefore accepted by patients.

However, in human-robot interactions, acceptance must be widely established before new robotic solutions can be successfully integrated into everyday life and the medical field. One reason for this is that, in contrast to most other humans, humans do not inherently accept machines and robots. Hence, it is essential to analyze human-robot interactions and understand the different factors contributing to acceptance.

Therefore, the last chapter of this thesis focuses specifically on acceptance in medical robotic systems, which involves interaction between the robot and the patient.

9.2 Related Work

Medical robotics itself is a vast research topic. However, the research on acceptance in the medical field is limited. While translational research aims to transfer laboratory-based projects to their application environments, it only briefly touches on acceptance.

Acceptance is currently mainly analyzed in the context of robotic surgery [147, 148, 149] and physician-robot interactions [150, 151]. Both cases focus on the acceptance and trust of the physician and surgical staff toward the robot. Attia et al. [152] introduced a framework for trusted autonomy in surgical robotics but only analyzed the robot-surgeon relationship. Torrent-Sellens et al. [153] investigated trust in robot-assisted surgeries in Europe. The study found that introducing surgical systems primarily depends on the patient's preferences. However, this preference is not based on direct interaction with the robot because patients are typically anesthetized during procedures that involve surgical robotic systems. Therefore, acceptance is limited to the general trust in the statistical outcome of such procedures. In contrast, acceptance plays a distinct role when the robot interacts with an awake human, as seen in elderly care settings [154, 155, 156, 157].

Little research has been performed in clinical and non-interventional settings. Bodenhagen et al. [158, 159] analyzed the impact of different interaction channels on trust in the robot. Results showed that verbal cues improved transparency, and transparency in itself has a positive impact on trust. Weigelin et al. [160] concluded that vocal interactions increase trust more than solely kinesthetic interactions.

9.3 Interaction Types

In typical scenarios, humans would like to interact jointly with the robot to achieve a task. Therefore, the research generally refers to this as a 'human-robot' interaction. During surgery, this could be rephrased as a 'surgeon-robot' interaction. However, diagnostic scenarios in which the patient is fully awake are distinctively different. In contrast to the above scenarios, the robotic system wishes to interact with the patient to achieve an imaging task. We, therefore, refer to these cases as 'robot-patient' interactions.

This thesis analyzes the acceptance of medical robotic imaging systems, particularly for tasks involving robot-patient interaction. Specifically, we evaluate the effect of different interaction and communication modes on acceptance during robotic ultrasound scanning. Communication is an exchange of information between individuals that can happen through different channels. Individuals can, for example, write notes, draw images or talk to each other to forward information. In our current setting, we focus on communication through speech, thereby, for example, transmitting information about the procedure or the robotic state from the robot to the patient. Furthermore, consent is communicated from the patient to the robot. During interactions, an action occurs between two individuals, or one reacts to the other. Gestural interaction could, for example, be a hand-waving motion to say goodbye to someone. Haptic interactions could be a handshake to greet someone or handing an object to someone. Generally, haptic interaction requires a closer distance between the involved individuals and

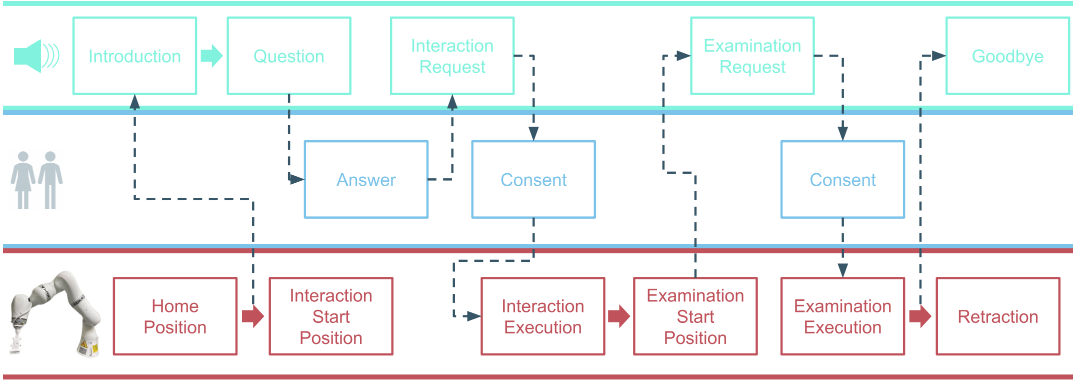


Fig. 9.1. Overview of the proposed framework, divided into verbal cues (top row), robot actions (bottom row), and human interaction (middle row). Dotted lines represent interactions between the three main parts. Adapted from [14].

can therefore be seen as a closer interaction than the gestural counterpart. Once humans accept gestural interactions with one another and see that these actions do not harm them, humans are willing to participate in haptic interactions. In this thesis, we, therefore, initially focus on gestural interactions. Once more acceptance is created, these can be extended to haptic interactions.

9.4 Robotic Interaction

To increase acceptance, we need to analyze which interaction parameters influence acceptance and how this influence is taking place. Therefore, we propose a general pipeline for applying and evaluating different interaction and communication modes (see Figure 9.1). The base pipeline allows gestural or haptic and visual interaction and communication between the robot and the human. Furthermore, the content of each interaction and communication block can be adapted for specific interaction scenarios.

Here, we propose a detailed interaction and communication plan for a robotic US scanning scenario and take inspiration from sonographers' everyday interactions with patients in routine clinical settings. The communication begins with the robot's self-introduction and asks the patient if they are familiar with the ultrasound scan procedure. This superficial conversation is intended to relax the patient and give the impression that the robot is taking the patient's response into account. The robot then performs an initial gestural interaction, suggesting a friendly no-touch high-five. Through this interaction, the robot shows awareness of action and position in space. Next, the robot asks to approach the neck for the US scan. After receiving consent, the robot approaches the neck, performs the scan, retracts, and says goodbye. Throughout the process, the robot continuously seeks consent and shows awareness of the patient's responses. Visual details from the approach are shown in Figure 9.2.

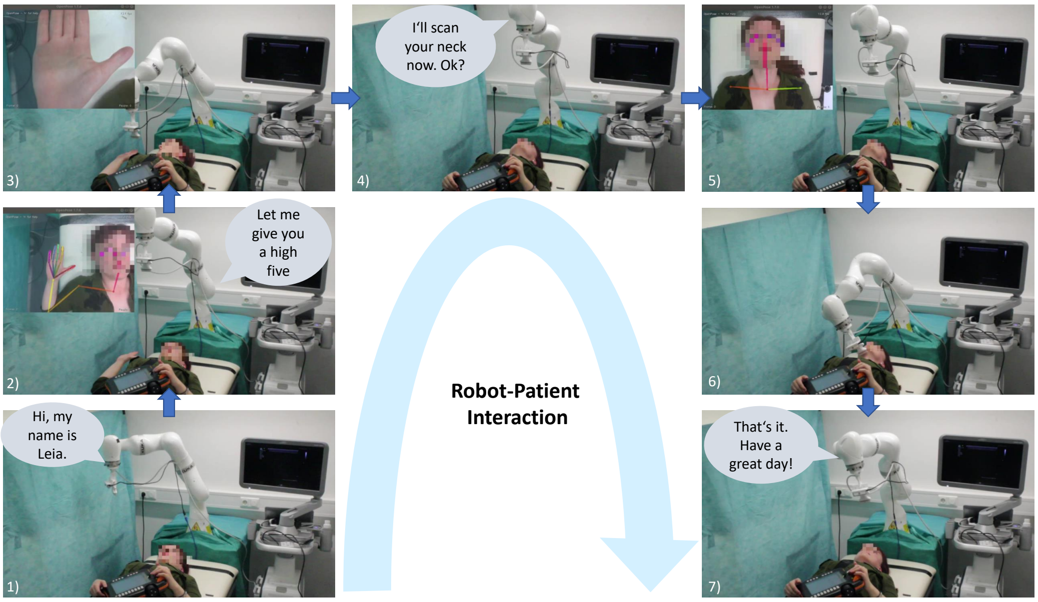


Fig. 9.2. Excerpts from the acceptance approach: 1) the robot introduces itself, 2) and 3) the robot gives a high five, 4) the robot asks for consent to approach the neck, 5) and 6) the robot approaches the neck, 7) the robot retracts and says goodbye. 2) and 5) show hand and neck key point detection via OpenPose.

9.5 Acceptance Evaluation

The concept of acceptance involves a person perceiving the actions performed by the robot as beneficial and non-harmful. Despite its straightforward definition, quantifying acceptance can be challenging due to its subjective nature, which can be influenced by factors such as previous experiences, environment, and more. In this thesis, we, therefore, propose two different ways to evaluate acceptance, qualitatively and quantitatively.

First, we present a questionnaire measuring a trust score, specifically for robotic ultrasound scanning. The questionnaire is based on work by Schaefer et al. [161]. In this original work, the authors introduced a set of 40 questions to evaluate trust in human-robot interactions. The scenarios were specified for long-term and collaborative interactions, such as in elderly care. For our case of robotic US scanning, we adapted this questionnaire. We chose 16 of the 40 available statements applicable to our setup (see Table 9.1). For example, statements related to collaborative tasks were not included. Furthermore, the original questionnaire evaluates long-term interactions. Therefore, the questions must be answered on a scale from $0 - 100\%$, depending on how often the robot fulfills the statement. We adapted this ranking to a 5-point Likert scale (see Eq. 9.1) because, in our case, each participant took part in each procedure only once.

1. The robot is responsible.
2. The robot is incompetent.
3. The robot is friendly.
4. The robot is reliable.
5. The robot is pleasant.
6. The robot is unresponsive.
7. The robot is autonomous.
8. The robot is predictable.
9. The robot acts consistently.
10. The robot functions successfully.
11. The robot malfunctions.
12. The robot communicates clearly.
13. The robot meets the needs of the task.
14. The robot provides appropriate information.
15. The robot communicates with people.
16. The robot performs exactly as instructed.

Tab. 9.1. Questionnaire statements, adapted from [161]. The users were asked to rank the statements in five categories from strongly agree to strongly disagree.

The trust score is then computed by assigning a value to each option such that

$$\text{trust score} = \begin{cases} 1 & \text{if strongly disagree} \\ 2 & \text{if disagree} \\ 3 & \text{if neutral} \\ 4 & \text{if agree} \\ 5 & \text{if strongly agree} \end{cases} \tag{9.1}$$

The results of Statements 2, 6, and 11 are inverted, respectively.

To evaluate acceptance more quantitatively, we also measured the heart rate. We assume that increased or unstable heart rates are signs of stress and therefore relate to a lower acceptance.

9.6 Experiments and Results

This section will look at the evaluation of acceptance through increased interaction. First, we will describe the study setup, and afterward, we will present the study results.

9.6.1 Experiments

The setup is similar to the one described in Chapter 5, consisting of a robotic manipulator (KUKA LBR iiwa, KUKA AG, Augsburg, Germany), a US machine (Siemens ACUSON Juniper, Siemens Healthineers, Germany), and a linear transducer, attached to the end-effector of the robotic arm. In addition, we attached an RGB-D camera (RealSense D435, Intel, USA) to the end-effector, and the user is provided with an emergency stop, stopping the robot motion immediately if needed. The workstation runs ROS, iiwa stack, and OpenPose [146, 162] for robot navigation. Real-time detection of the human body and facial and hand key points is achieved through OpenPose. A margin of two centimeters between the human key points and the goal US probe pose is introduced to adhere to safety aspects. We used the same approach

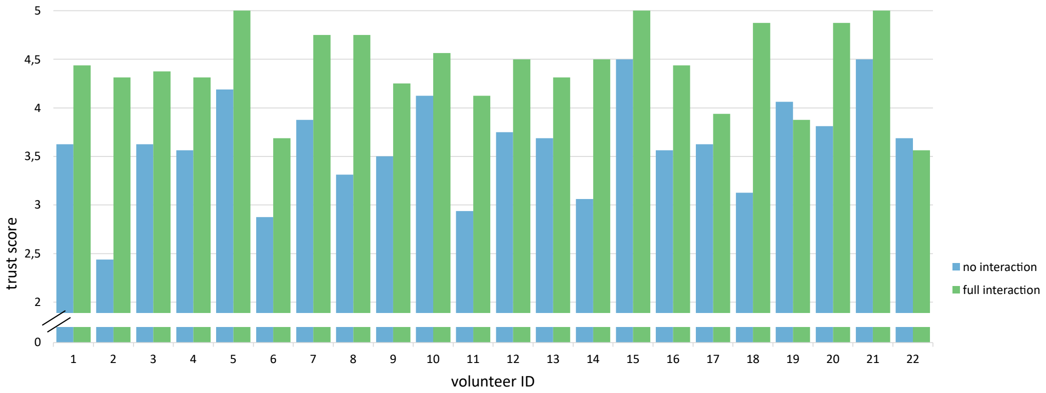


Fig. 9.3. Block diagram of the overall trust score for each participant in the no interaction (blue) and full interaction (green) case. Adapted from [14].

described in Section 8.6 to detect the neck target point. To execute the high five, we extracted three key points on the palm, forming a triangle. We then extracted the center point and normal vector from this triangle as the target position and orientation for the robot.

The evaluation consisted of the questionnaire mentioned above, filled by each participant directly after each approach, and the heart rate measurements of each participant. For the latter evaluation, we attached a heart rate monitor (HRM-Dual, Garmin, USA) to the participant's chest. The heart rate measurements during the robotic approach to the neck were compared in both settings.

For the user study, we recruited a sample of 20 individuals. The participants were a mixture of medical and computer science researchers and comprised of ten men and ten women, with an age range of $20 - 33$ years and a mean age of 25.8 ± 2.62 years. None of the participants were familiar with the system, and their prior experience with robotic applications varied. 14 participants had low experience, four had medium experience, and two had high prior robotic experience. The participants were randomly assigned to one of the two orders of approach options. Furthermore, they were informed about the aim and procedure of the study. No prior training was required.

9.6.2 Results

The average trust score for all participants was 3.61 ± 0.50 and 4.43 ± 0.40 for the scenarios with no interaction and full interaction, respectively. The higher trust score for the full interaction case can also be observed in Figure 9.3, which shows each participant's trust score in a block diagram. In 19 out of 20 cases, full interaction resulted in a higher trust score. Figure 9.4 shows the average trust score for each questionnaire statement. Statements 3, 8, 12, 14, and 15 show the most significant trust score difference between both approaches, whereas statements 1, 2, 4, 9, and 10 show the smallest differences. Furthermore, statement 15 shows the lowest trust score rating overall statements.

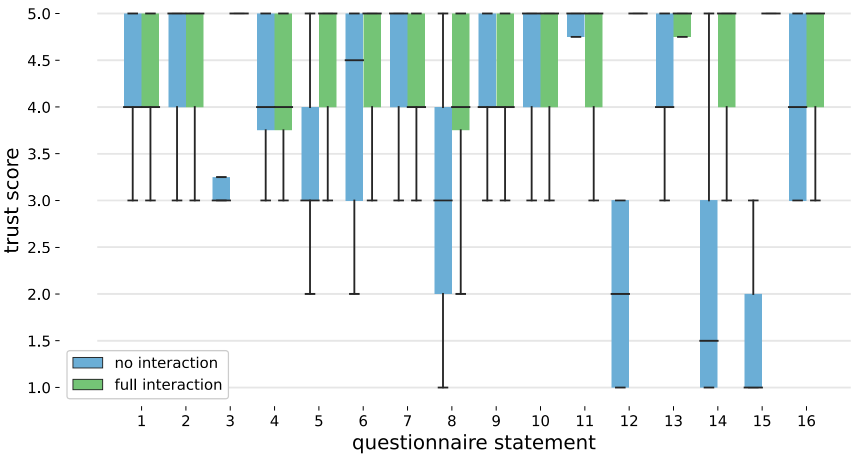


Fig. 9.4. Box diagram of the trust score for each questionnaire statement for the no interaction (blue) and full interaction case (green).Adapted from [14].

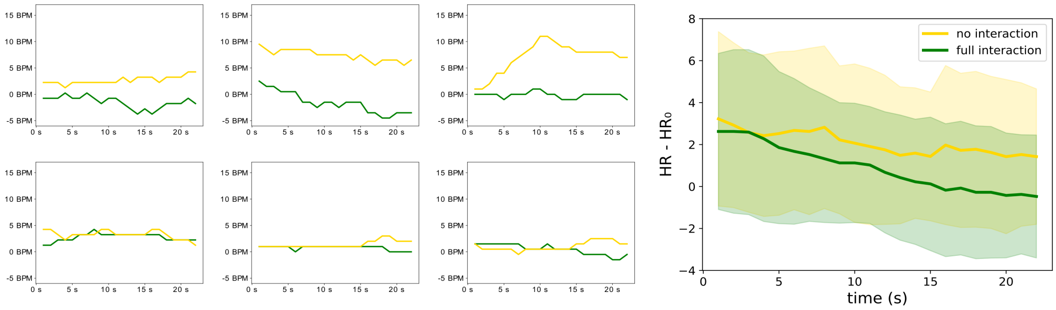


Fig. 9.5. Comparison between the resting heart rate (HR0) and the heart rate during the procedure (HR). Right: Mean and standard deviation of the heart rate difference between both states over all participants. Left: heart rate difference examples from individual participants, for which the full interaction is favorable (top row) and for which both options are favorable (bottom row). Reproduced from [14].

Their heart rate values were analyzed to assess the participants' stress levels objectively. The median heart rate in a resting state was calculated as a baseline for each participant. Then the difference between the median and the heart rate during the procedure was computed. The average result of this comparison across all participants is shown in Figure 9.5 on the right. Furthermore, we depict heart rate evolutions during the procedure for different individuals. In our results, the heart rate difference compared to the rest state showed a decrease of 1.8 BPM in the no interaction condition and a more substantial decrease of 3.1 BPM in the full interaction condition. Analyzing the heart rate changes per individual showed mixed results, favoring the full interaction case in some participants and both approaches in others.

9.7 Discussion

The evaluation of the proposed pipeline led to different insights, which will be discussed here. First, we will discuss the questionnaire results. Next, we will analyze the heart rate measurements and debate the usefulness of heart rate measurements to measure acceptance. Lastly, we will look at the influence of the current study cohort on the results.

9.7.1 Questionnaire Analysis

The questionnaire results indicate a positive impact of human-like interactions and communication on the trust score and, therefore, on acceptance. In statement 11, the no interaction scenario slightly outperforms the full interaction scenario. This difference can be attributed to two outliers in the questionnaire responses, where the robot's motion was not typical, leading to a decrease in trust in the robot's ability to function effectively. Statements 1, 2, 4, 7, 9, and 10 show the smallest differences between both approaches. 'Incompetent', 'reliable', 'act consistently', and 'function successfully' are all related to the direct task of the robot, and this task did not change between both approaches, possibly explaining the similarity in trust scores. Statement 1 concerns the responsibility of the robot. In the current setup, the participants might not have experienced a task requiring high responsibility and therefore rated both approaches similarly. Statement 7 evaluates the perception of the system's autonomy. The results indicate that introducing interactions and consent through communication can lead to a decrease in the perceived autonomy of the system.

In contrast, statements 3, 8, 12, 14, and 15 show an increased trust score for the full interaction case. The last three statements are related to communication, showing that an interaction's communicative aspect can increase acceptance. Statement 8 evaluates the predictability of the system, and the increased trust score suggests that through communication and initial interactions, the perceived predictability of the robotic system can be increased. Statement 3 assesses the friendliness of the system and reflects the emotional aspect of acceptance. The results indicate that the proposed interaction and communication positively influence this aspect.

While the average trust scores for the full interaction scenario of statements 5, 8, 9, 12, and 16 were high (ranging from 4.1 to 4.75), these statements also included outliers with a low trust score of 2. This finding could indicate that the perception of different interactions and communications is subjective. Additionally, participant feedback suggests a preference for communication over a simple high-five gesture.

9.7.2 Heart Rate Analysis

We obtained heart rate data to quantitatively assess interactions and communications' effect on patients' acceptance. We analyzed the change in heart rates relative to resting heart rates, indicating a change in stress level before and during the procedure. The results show a slightly greater decrease in heart rates in the full interaction scenario, which may suggest a marginal reduction in stress levels during the procedure. However, analyzing the individual results mostly show only a small difference in heart rates, making it challenging to choose a more stress-reducing scenario for each participant. Overall, the inclusion of heart rate analysis provides an alternative measurement to the questionnaire. The results marginally support the questionnaire findings, as a reduction in stress can also be associated with higher acceptance. However, heart rate results are not significant and do not show a clear result if analyzed alone.

9.7.3 Study Population

The current user study involved young and healthy volunteers. However, this cohort does not accurately reflect patient populations undergoing thyroid diagnostics. These groups are generally older and less familiar with modern technology. Therefore, further studies with patient populations should be conducted to gain more comprehensive and realistic insights into the impact of interactions and communication. These studies could also lead to the development of patient- and examination-specific interactions. Additionally, a larger sample size would enable the exploration of various factors, such as the impact of different communication methods, voice types, and verbal cues.

9.8 Future Work

Analyzing acceptance in medical robotic imaging can be advanced in several areas based on insights from the results presented here. The heart rate monitoring showed supportive results. However, it would be helpful to analyze other qualitative metrics that could indicate the degree of acceptance of these systems. Such a metric would allow for analyzing acceptance in more detail and on a more subconscious level compared to the current questionnaire. The current questionnaire is also an adaption of a questionnaire meant to analyze trust in collaborative and long-term interactions. Future works could design questionnaires for the specific case of medical robotic imaging to analyze acceptance in a more tailored way.

More extensive and complex communication modes can be introduced to mimic sonographers' interaction with the patient. These modes could be enabled by receiving the patient's name through the patient report and directly addressing the patient by their name. Furthermore, the planned procedure can be described in more detail to increase the knowledge transfer between the robotic system and the patient.

Subsequent studies should be performed with cohorts that resemble patient groups and are significantly bigger. These studies would then allow the evaluation of more patient-specific interactions and communication. Moreover, bigger sample sizes would enable an individual analysis of single interaction and communication blocks. The content of communication or the type of voice of the robotic system could, for example, be evaluated in detail.

In this last contribution part, we analyzed acceptance in medical robotic systems. So now that we proposed methods for intelligent, applicable, and acceptable robotic US scanning, we will conclude this thesis in the following last part.

Part VII

Conclusion

Conclusion

10

In this thesis, we proposed methods to jointly enhance the three named pillars of translational research on robotic ultrasound: intelligence, application, and acceptance (see Figure 10.1).

We introduced physics-inspired augmentation techniques to produce variable high-quality US data. Furthermore, we proposed a method for robotically acquiring US compoundings by mimicking a sonographer's workflow. Specifically, we introduced a framework for general volume coverage and acoustic shadow reduction by rescanning low-quality areas. Both methods produce valuable US data and increase the intelligence of robotic US systems.

Next, we made intelligent robotic ultrasound systems applicable by introducing a robotic framework for thyroid volumetry. In this case, we not only looked at the robotic US scanning itself but combined it with an automatic thyroid segmentation. Comparing our method to the current gold standard in clinical practice, we reduced variability, increased accuracy, and became independent from clinical experts during the acquisition process. Therefore, we were able to highlight the possible impact of these robotic systems.

Lastly, we analyzed patient acceptance of robotic US systems. To achieve this task, we mimicked a human sonographer's interactions and communications with the patient. Through this approach, we showed that robot-patient interactions and communication could increase acceptance of the robotic system. These findings can be used to accelerate further research in patient acceptance of robotic medical imaging systems. Once patient acceptance of robotic systems has been broadly introduced, these systems can be applied in clinical practice and further developed to enable trust in the systems.

Overall, this thesis presented a roadmap for translational research on robotic ultrasound. The mentioned research pillars are often followed and explored as a standalone task but are rarely combined. However, the joint analysis of all three pillars is the key to translation from research to clinical practice - from bench to bedside. After such a successful translation, robotic US has the potential to become the new standard in medical imaging technology with optimized workflows comparable to the acquisition of MRI or CT. Patients will be sent to the robotic US acquisition suit, and technicians will be there to guide them through the process, thus saving the doctors valuable time. Afterward, doctors make diagnostic and treatment decisions based on standardized robotic US scans and additional image analysis support from the robotic system. Therefore, by following the roadmap presented in this thesis, future research can simultaneously implement and evaluate robotic ultrasound systems on the topics of intelligence, application, and acceptance. Thereby paving the way for robotic ultrasound systems that can be integrated into clinical practice.

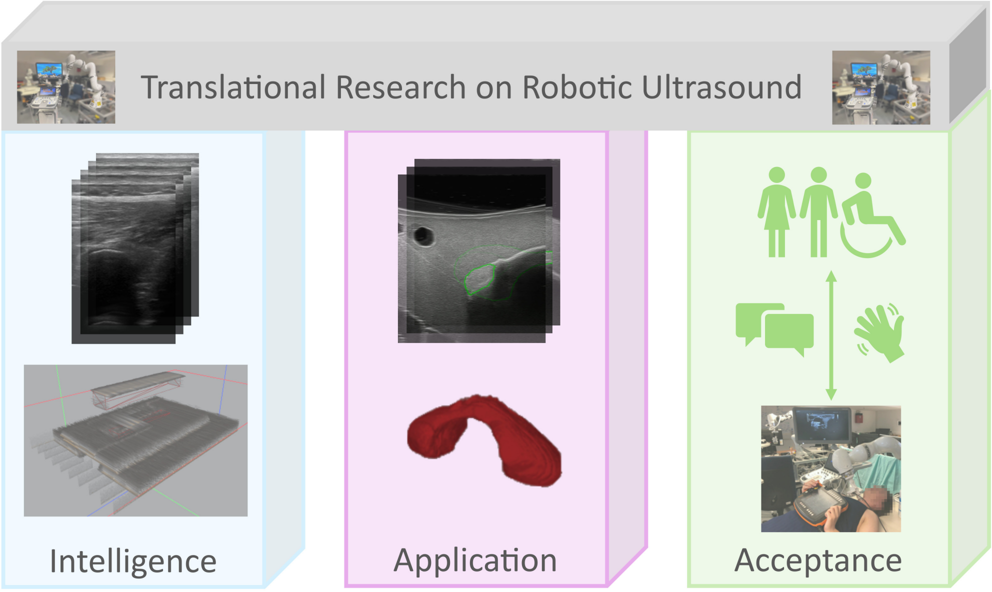


Fig. 10.1. The three pillars, intelligence, application, and acceptance, required for successful translation research on robotic ultrasound, filled with proposed methods to enable and enhance all three research topics.

Part VIII

Appendix

A

Authored and Co-authored Publications

Authored

1. M. Tirindelli* **C. Eilers***, W. A. Simson, M. Paschali, M. F. Azampour, N. Navab. *“Rethinking Ultrasound Augmentation: A Physics-Inspired Approach.”*, Medical Image Computing and Computer Assisted Intervention (MICCAI), 2021 [11] (*authors contributed equally) **[Poster Presentation].**

2. M. Krönke*, **C. Eilers***, D. Dimova, M. Köhler, G. Buschner, L. Schweiger, L. Konstantinidou, M. Makowski, J. Nagarajah, N. Navab, W. Weber, T. Wendler. *“Tracked 3D Ultrasound and Deep Neural Network-based Thyroid Segmentation reduce Interobserver Variability in Thyroid Volumetry.”* PLOS ONE, 2022 [140] (*authors contributed equally)

3. J. Zielke*, **C. Eilers***, B. Busam, W. Weber, N. Navab, T. Wendler. *“RSV: Robotic Sonography for Thyroid Volumetry.”* IEEE Robotics and Automation Letters, 2022 [13] (*authors contributed equally). **[Oral and Poster Presentation at ICRA2023].**

4. V. Sutedjo*, M. Tirindelli*, **C. Eilers***, W. Simson, B. Busam, N. Navab. *“Acoustic Shadowing Aware Robotic Ultrasound: Lighting up the Dark.”* IEEE Robotics and Automation Letters, 2022 [12] (*authors contributed equally).

5. **C. Eilers***, R. van Kemenade*, B. Busam, N. Navab. *“On the Importance of Patient Acceptance for Medical Robotic Imaging.”* [arXiv preprint], arXiv:2302.06208, 2023. [14] (*authors contributed equally).

Co-authored

1. M. Wysocki, M. F. Azampour, **C. Eilers**, B. Busam, M. Salehi, N. Navab. *“Ultra-NeRF: Neural Radiance Fields for Ultrasound Imaging.”* arXiv.2301.10520.

B Abstracts of Publications not Discussed in this Thesis

Tracked 3D Ultrasound and Deep Neural Network-based Thyroid Segmentation reduce Interobserver Variability in Thyroid Volumetry

M. Krönke, **C. Eilers**, D. Dimova, M. Köhler, G. Buschner, L. Schweiger, L. Konstantinidou, M. Makowski, J. Nagarajah, N. Navab, W. Weber, T. Wendler. PLOS ONE, 2022

Abstract

Thyroid volumetry is crucial in the diagnosis, treatment, and monitoring of thyroid diseases. However, conventional thyroid volumetry with 2D ultrasound is highly operator-dependent. This study compares 2D and tracked 3D ultrasound with an automatic thyroid segmentation based on a deep neural network regarding inter- and intraobserver variability, time, and accuracy. Volume reference was MRI. 28 healthy volunteers (24 - 50 a) were scanned with 2D and 3D ultrasound (and by MRI) by three physicians (MD 1, 2, 3) with different experience levels (6, 4, and 1 a). In the 2D scans, the thyroid lobe volumes were calculated with the ellipsoid formula. A convolutional deep neural network (CNN) automatically segmented the 3D thyroid lobes. 26, 6, and 6 random lobe scans were used for training, validation, and testing, respectively. On MRI (T1 VIBE sequence) the thyroid was manually segmented by an experienced MD. MRI thyroid volumes ranged from 2.8 to $16.7ml$ (mean 7.4, SD 3.05). The CNN was trained to obtain an average Dice score of 0.94. The interobserver variability comparing two MDs showed mean differences for 2D and 3D respectively of 0.58 to $0.52ml$ (MD1 vs. 2), -1.33 to $-0.17ml$ (MD1 vs. 3) and -1.89 to $-0.70ml$ (MD2 vs. 3). Paired samples t-tests showed significant differences for 2D ($p = .140$, $p = .002$ and $p = .002$) and none for 3D ($p = .176$, $p = .722$ and $p = .057$). Intraobsever variability was similar for 2D and 3D ultrasound. Comparison of ultrasound volumes and MRI volumes showed a significant difference for the 2D volumetry of all MDs ($p = .002$, $p = .009$, $p < .001$), and no significant difference for 3D ultrasound ($p = .292$, $p = .686$, $p = 0.091$). Acquisition time was significantly shorter for 3D ultrasound. Tracked 3D ultrasound combined with a CNN segmentation significantly reduces interobserver variability in thyroid volumetry and increases the accuracy of the measurements with shorter acquisition times.

Bibliography

[1] V. Chan and A. Perlas. “Basics of ultrasound imaging”. In: *Atlas of ultrasound-guided procedures in interventional pain management* (2011), pp. 13–19 (cit. on p. 3).

[2] R. H. Hashemi, W. G. Bradley, and C. J. Lisanti. *MRI: the basics: The Basics*. Lippincott Williams & Wilkins, 2012 (cit. on p. 3).

[3] J. T. Hathcock and R. L. Stickle. “Principles and concepts of computed tomography”. In: *Veterinary Clinics of North America: Small Animal Practice* 23.2 (1993), pp. 399–415 (cit. on p. 3).

[4] R. Kojcev, A. Khakzar, B. Fuerst, et al. “On the reproducibility of expert-operated and robotic ultrasound acquisitions”. In: *International Journal of Computer Assisted Radiology and Surgery* 12 (Mar. 2017) (cit. on pp. 3, 23, 43, 60, 62, 71).

[5] P. Wareluk and W. Jakubowski. “Evaluation of musculoskeletal symptoms among physicians performing ultrasound”. In: *Journal of Ultrasonography* 17.70 (2017), pp. 154–159 (cit. on p. 3).

[6] K. Evans, S. Roll, and J. Baker. “Work-related musculoskeletal disorders (WRMSD) among registered diagnostic medical sonographers and vascular technologists: a representative sample”. In: *Journal of Diagnostic Medical Sonography* 25.6 (2009), pp. 287–299 (cit. on p. 3).

[7] G. Harrison and A. Harris. “Work-related musculoskeletal disorders in ultrasound: Can you reduce risk?” In: *Ultrasound* 23.4 (2015), pp. 224–230 (cit. on p. 3).

[8] J. Korenblit, D. M. Tholey, J. Tolin, et al. “Effect of the time of day and queue position in the endoscopic schedule on the performance characteristics of endoscopic ultrasound-guided fine-needle aspiration for diagnosing pancreatic malignancies”. In: *Endoscopic Ultrasound* 5.2 (2016), p. 78 (cit. on p. 3).

[9] A. Fenster and D. B. Downey. “3-D ultrasound imaging: A review”. In: *IEEE Engineering in Medicine and Biology magazine* 15.6 (1996), pp. 41–51 (cit. on p. 3).

[10] B. Busam, M. Esposito, B. Frisch, and N. Navab. “Quaternionic upsampling: Hyperspherical techniques for 6 dof pose tracking”. In: *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE. 2016, pp. 629–638 (cit. on pp. 3, 14, 66).

[11] M. Tirindelli, C. Eilers, W. Simson, M. Paschali, M. F. Azampour, and N. Navab. “Rethinking ultrasound augmentation: A physics-inspired approach”. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24*. Springer. 2021, pp. 690–700 (cit. on pp. 5, 33, 35–39, 93).

[12] V. Sutedjo, M. Tirindelli, C. Eilers, W. Simson, B. Busam, and N. Navab. “Acoustic Shadowing Aware Robotic Ultrasound: Lighting up the Dark”. In: *IEEE Robotics and Automation Letters* 7.2 (2022), pp. 1808–1815 (cit. on pp. 6, 45, 50–56, 93).

[13] J. Zielke, C. Eilers, B. Busam, W. Weber, N. Navab, and T. Wendler. “RSV: robotic sonography for thyroid volumetry”. In: *IEEE Robotics and Automation Letters* 7.2 (2022), pp. 3342–3348 (cit. on pp. 6, 64, 65, 69, 70, 93).

[14] C. Eilers, R. van Kemenade, B. Busam, and N. Navab. *On the Importance of Patient Acceptance for Medical Robotic Imaging*. 2023. arXiv: `2302.06208` (cit. on pp. 6, 79, 82, 83, 93).

[15] L. Demi. “Practical guide to ultrasound beam forming: Beam pattern and image reconstruction analysis”. In: *Applied Sciences* 8.9 (2018), p. 1544 (cit. on p. 13).

[16] A. C. Luchies and B. C. Byram. “Deep neural networks for ultrasound beamforming”. In: *IEEE transactions on medical imaging* 37.9 (2018), pp. 2010–2021 (cit. on p. 13).

[17] R. Prevost, M. Salehi, S. Jagoda, et al. “3D freehand ultrasound without external tracking using deep learning”. In: *Medical image analysis* 48 (2018), pp. 187–202 (cit. on p. 14).

[18] M. Luo, X. Yang, H. Wang, L. Du, and D. Ni. “Deep Motion Network for Freehand 3D Ultrasound Reconstruction”. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part IV*. Springer. 2022, pp. 290–299 (cit. on p. 14).

[19] Y. Kwoh, J. Hou, E. Jonckheere, and S. Hayati. “A robot with improved absolute positioning accuracy for CT guided stereotactic brain surgery”. In: *IEEE Transactions on Biomedical Engineering* 35.2 (1988), pp. 153–160 (cit. on p. 17).

[20] C. Hennersperger, B. Fuerst, S. Virga, et al. “Towards MRI-based autonomous robotic US acquisitions: a first feasibility study”. In: *IEEE transactions on medical imaging* 36.2 (2017), pp. 538–548 (cit. on pp. 18, 27, 68).

[21] D. Arikan. “Anatomy-dependent force adjustment for robotic ultrasound through external force sensing”. Bachelor’s Thesis. Technical University of Munich, 2022 (cit. on p. 18).

[22] M. E. Moran. “Evolution of robotic arms”. In: *Journal of robotic surgery* 1.2 (2007), pp. 103–111 (cit. on p. 18).

[23] M. Ceccarelli. “Fundamentals of the mechanics of robots”. In: *Fundamentals of Mechanics of Robotic Manipulation*. Dordrecht: Springer Netherlands, 2004, pp. 73–240 (cit. on p. 18).

[24] J. M. Selig. *Geometric fundamentals of robotics*. Vol. 128. Springer, 2005 (cit. on p. 18).

[25] D. D. Ardayfio. *Fundamentals of robotics*. CRC Press, 2020 (cit. on p. 18).

[26] S. B. Niku. *Introduction to robotics: analysis, control, applications*. John Wiley & Sons, 2020 (cit. on p. 18).

[27] J. M. Beer, A. D. Fisk, and W. A. Rogers. “Toward a framework for levels of robot autonomy in human-robot interaction”. In: *Journal of human-robot interaction* 3.2 (2014), p. 74 (cit. on p. 20).

[28] G.-Z. Yang, J. Cambias, K. Cleary, et al. *Medical robotics—Regulatory, ethical, and legal considerations for increasing levels of autonomy*. 2017 (cit. on p. 20).

[29] D. R. Swerdlow, K. Cleary, E. Wilson, B. Azizi-Koutenaei, and R. Monfaredi. “Robotic arm–assisted sonography: Review of technical developments and potential clinical applications”. In: *American Journal of Roentgenology* 208.4 (2017), pp. 733–738 (cit. on p. 21).

[30] A. M. Priester, S. Natarajan, and M. O. Culjat. “Robotic ultrasound systems in medicine”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 60.3 (2013), pp. 507–523 (cit. on p. 21).

[31] F. von Haxthausen, S. Böttger, D. Wulff, J. Hagenah, V. García-Vázquez, and S. Ipsen. “Medical robotics for ultrasound imaging: current systems and future trends”. In: *Current Robotics Reports* 2.1 (2021), pp. 55–71 (cit. on p. 21).

[32] K. Li, Y. Xu, and M. Q.-H. Meng. “An Overview of Systems and Techniques for Autonomous Robotic Ultrasound Acquisitions”. In: *IEEE Transactions on Medical Robotics and Bionics* 3.2 (2021), pp. 510–524 (cit. on pp. 21, 22).

[33] A. M. Mikula-Curtis, J. L. Marshall, and L. M. Bruno. *Ultrasound imaging system with touch-pad pointing device*. US Patent 6,135,958. Oct. 2000 (cit. on p. 21).

[34] A. V. Gonzales, P. Cinquin, J. Troccaz, et al. “TER: a system for robotic tele-echography”. In: *Lecture Notes in Computer Science* (2001), pp. 326–334 (cit. on p. 21).

[35] A. Vilchis, J. Troccaz, P. Cinquin, K. Masuda, and F. Pellissier. “A new robot architecture for tele-echography”. In: *IEEE Transactions on Robotics and Automation* 19.5 (2003), pp. 922–926 (cit. on p. 21).

[36] T. Essomba, L. Nouaille, M. Laribi, G. Poisson, and S. Zeghloul. “Design process of a robotized tele-echography system”. In: *Applied Mechanics and Materials*. Vol. 162. Trans Tech Publ. 2012, pp. 384–393 (cit. on p. 21).

[37] F. Pierrot, E. Dombre, E. Dégoulange, et al. “Hippocrate: A safe robot arm for medical applications with force feedback”. In: *Medical Image Analysis* 3.3 (1999), pp. 285–300 (cit. on p. 21).

[38] M.-A. Janvier, L.-G. Durand, M.-H. R. Cardinal, et al. “Performance evaluation of a medical robotic 3D-ultrasound imaging system”. In: *Medical image analysis* 12.3 (2008), pp. 275–290 (cit. on p. 21).

[39] G. P. Mylonas, P. Giataganas, M. Chaudery, V. Vitiello, A. Darzi, and G.-Z. Yang. “Autonomous eFAST ultrasound scanning by a robotic manipulator using learning from demonstrations”. In: *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2013, pp. 3251–3256 (cit. on p. 21).

[40] M. Victorova, D. Navarro-Alarcon, and Y.-P. Zheng. “3d ultrasound imaging of scoliosis with force-sensitive robotic scanning”. In: *2019 third IEEE international conference on robotic computing (IRC)*. IEEE. 2019, pp. 262–265 (cit. on p. 21).

[41] S. E. Salcudean, G. Bell, S. Bachmann, W.-H. Zhu, P. Abolmaesumi, and P. D. Lawrence. “Robot-assisted diagnostic ultrasound–design and feasibility experiments”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI’99: Second International Conference, Cambridge, UK, September 19-22, 1999. Proceedings 2*. Springer. 1999, pp. 1062–1071 (cit. on p. 21).

[42] S. E. Salcudean, W. H. Zhu, P. Abolmaesumi, S. Bachmann, and P. D. Lawrence. “A robot system for medical ultrasound”. In: *Robotics Research: The Ninth International Symposium*. Springer. 2000, pp. 195–202 (cit. on p. 21).

[43] P. Abolmaesumi, S. Salcudean, and W. Zhu. “Visual servoing for robot-assisted diagnostic ultrasound”. In: *Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Cat. No. 00CH37143)*. Vol. 4. IEEE. 2000, pp. 2532–2535 (cit. on p. 21).

[44] P. Abolmaesumi, M. R. Sirouspour, and S. E. Salcudean. “Real-time extraction of carotid artery contours from ultrasound images”. In: *Proceedings 13th IEEE Symposium on Computer-Based Medical Systems. CBMS 2000*. IEEE. 2000, pp. 181–186 (cit. on p. 21).

[45] P. Abolmaesumi, S. E. Salcudean, W.-H. Zhu, M. R. Sirouspour, and S. P. DiMaio. “Image-guided control of a robot for medical ultrasound”. In: *IEEE transactions on robotics and automation* 18.1 (2002), pp. 11–23 (cit. on p. 21).

[46] H. T. Sen, M. A. L. Bell, I. Iordachita, J. Wong, and P. Kazanzides. “A cooperatively controlled robot for ultrasound monitoring of radiation therapy”. In: *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2013, pp. 3071–3076 (cit. on p. 21).

[47] T.-Y. Fang, H. K. Zhang, R. Finocchi, R. H. Taylor, and E. M. Boctor. “Force-assisted ultrasound imaging system through dual force sensing and admittance robot control”. In: *International journal of computer assisted radiology and surgery* 12 (2017), pp. 983–991 (cit. on p. 21).

[48] P. Arbeille, J. Ruiz, P. Herve, M. Chevillot, G. Poisson, and F. Perrotin. “Fetal tele-echography using a robotic arm and a satellite link”. In: *Ultrasound in Obstetrics and Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology* 26.3 (2005), pp. 221–226 (cit. on p. 21).

[49] P. Arbeille, G. Poisson, P. Vieyres, J. Ayoub, M. Porcher, and J. L. Boulay. “Echographic examination in isolated sites controlled from an expert center using a 2-D echograph guided by a teleoperated robotic arm”. In: *Ultrasound in medicine & biology* 29.7 (2003), pp. 993–1000 (cit. on p. 21).

[50] A. Krupa and F. Chaumette. “Control of an ultrasound probe by adaptive visual servoing”. In: *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2005, pp. 2681–2686 (cit. on pp. 21, 22).

[51] R. Mebarki, A. Krupa, and F. Chaumette. “2-D ultrasound probe complete guidance by visual servoing using image moments”. In: *IEEE Transactions on Robotics* 26.2 (2010), pp. 296–306 (cit. on p. 21).

[52] P. Chatelain, A. Krupa, and N. Navab. “Confidence-driven control of an ultrasound probe”. In: *IEEE Transactions on Robotics* 33.6 (2017), pp. 1410–1424 (cit. on p. 21).

[53] A. Gourdon, P. Poignet, G. Poisson, P. Vieyres, and P. Marche. “A new robotic mechanism for medical application”. In: *1999 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (Cat. No. 99TH8399)*. IEEE. 1999, pp. 33–38 (cit. on p. 21).

[54] A. Krupa, G. Fichtinger, and G. D. Hager. “Real-time tissue tracking with B-mode ultrasound using speckle and visual servoing”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007: 10th International Conference, Brisbane, Australia, October 29-November 2, 2007, Proceedings, Part II 10*. Springer. 2007, pp. 1–8 (cit. on p. 22).

[55] C. Nadeau and A. Krupa. “A multi-plane approach for ultrasound visual servoing: application to a registration task”. In: *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2010, pp. 5706–5711 (cit. on p. 22).

[56] C. Nadeau, A. Krupa, J. Petr, and C. Barillot. “Moments-based ultrasound visual servoing: From a mono-to multiplane approach”. In: *IEEE Transactions on Robotics* 32.6 (2016), pp. 1558–1564 (cit. on p. 22).

[57] R. Nakadate, J. Solis, A. Takanishi, M. Sugawara, K. Niki, and E. Minagawa. “Development of the ultrasound probe holding robot WTA-1RII and an automated scanning algorithm based on ultrasound image feedback”. In: *ROMANSY 18 Robot Design, Dynamics and Control: Proceedings of The Eighteenth CISM-IFToMM Symposium*. Springer. 2010, pp. 359–366 (cit. on p. 22).

[58] R. Nakadate, J. Solis, A. Takanishi, E. Minagawa, M. Sugawara, and K. Niki. “Out-of-plane visual servoing method for tracking the carotid artery with a robot-assisted ultrasound diagnostic system”. In: *2011 IEEE International Conference on Robotics and Automation*. IEEE. 2011, pp. 5267–5272 (cit. on p. 22).

[59] A. S. B. Mustafa, T. Ishii, Y. Matsunaga, et al. “Development of robotic system for autonomous liver screening using ultrasound scanning device”. In: *2013 IEEE international conference on robotics and biomimetics (ROBIO)*. IEEE. 2013, pp. 804–809 (cit. on p. 22).

[60] C. Pahl and E. Supriyanto. “Design of automatic transabdominal ultrasound imaging system”. In: *2015 20th International Conference on Methods and Models in Automation and Robotics (MMAR)*. IEEE. 2015, pp. 435–440 (cit. on p. 22).

[61] R. Tsumura and H. Iwata. “Robotic fetal ultrasonography platform with a passive scan mechanism”. In: *International Journal of Computer Assisted Radiology and Surgery* 15 (2020), pp. 1323–1333 (cit. on p. 22).

[62] M. K. Welleweerd, A. G. de Groot, S. de Looijer, F. J. Siepel, and S. Stramigioli. “Automated robotic breast ultrasound acquisition using ultrasound feedback”. In: *2020 IEEE international conference on robotics and automation (ICRA)*. IEEE. 2020, pp. 9946–9952 (cit. on p. 22).

[63] S. Merouche, L. Allard, E. Montagnon, G. Soulez, P. Bigras, and G. Cloutier. “A robotic ultrasound scanner for automatic vessel tracking and three-dimensional reconstruction of b-mode images”. In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 63.1 (2015), pp. 35–46 (cit. on p. 23).

[64] S. Virga, O. Zettinig, M. Esposito, et al. “Automatic force-compliant robotic ultrasound screening of abdominal aortic aneurysms”. In: *2016 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE. 2016, pp. 508–513 (cit. on p. 23).

[65] C. Hennersperger, B. Fuerst, S. Virga, et al. “Towards MRI-Based Autonomous Robotic US Acquisitions: A First Feasibility Study”. In: *IEEE Transactions on Medical Imaging* 36.2 (2017), pp. 538–548 (cit. on pp. 23, 43).

[66] C. Graumann, B. Fuerst, C. Hennersperger, F. Bork, and N. Navab. “Robotic ultrasound trajectory planning for volume of interest coverage”. In: *2016 IEEE international conference on robotics and automation (ICRA)*. IEEE. 2016, pp. 736–741 (cit. on p. 23).

[67] R. Göbl, S. Virga, J. Rackerseder, B. Frisch, N. Navab, and C. Hennersperger. “Acoustic window planning for ultrasound acquisition”. In: *Int. J. Comput. Assist. Radiol. Surg.* 12.6 (2017), pp. 993–1001 (cit. on pp. 23, 43).

[68] Q. Huang, B. Wu, J. Lan, and X. Li. “Fully Automatic Three-Dimensional Ultrasound Imaging Based on Conventional B-Scan”. In: *IEEE Transactions on Biomedical Circuits and Systems* 12.2 (2018), pp. 426–436 (cit. on pp. 23, 62).

[69] Q. Huang, J. Lan, and X. Li. “Robotic Arm Based Automatic Ultrasound Scanning for Three-Dimensional Imaging”. In: *IEEE Transactions on Industrial Informatics* 15.2 (2019), pp. 1173–1182 (cit. on pp. 23, 62).

[70] J. T. Kaminski, K. Rafatzand, and H. K. Zhang. “Feasibility of Robot-Assisted Ultrasound Imaging with Force Feedback for Assessment of Thyroid Diseases”. In: *Proceedings of SPIE–the International Society for Optical Engineering* (2020) (cit. on pp. 23, 62).

[71] F. von Haxthausen, J. Hagenah, M. Kaschwich, M. Kleemann, V. Garcıa-Vázquez, and F. Ernst. “Robotized ultrasound imaging of the peripheral arteries–a phantom study”. In: *Current Directions in Biomedical Engineering* 6.1 (2020) (cit. on p. 23).

[72] Z. Jiang, H. Wang, Z. Li, et al. “Motion-aware robotic 3D ultrasound”. In: *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2021, pp. 12494–12500 (cit. on p. 23).

[73] Z. Jiang, Y. Zhou, Y. Bi, M. Zhou, T. Wendler, and N. Navab. “Deformation-aware robotic 3D ultrasound”. In: *IEEE Robotics and Automation Letters* 6.4 (2021), pp. 7675–7682 (cit. on p. 23).

[74] H. Hase, M. F. Azampour, M. Tirindelli, et al. “Ultrasound-guided robotic navigation with deep reinforcement learning”. In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2020, pp. 5534–5541 (cit. on pp. 23, 31).

[75] Y. Bi, Z. Jiang, Y. Gao, T. Wendler, A. Karlas, and N. Navab. “VesNet-RL: Simulation-based reinforcement learning for real-world us probe navigation”. In: *IEEE Robotics and Automation Letters* 7.3 (2022), pp. 6638–6645 (cit. on p. 23).

[76] G. Megali, O. Tonet, C. Stefanini, et al. “A computer-assisted robotic ultrasound-guided biopsy system for video-assisted surgery”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2001: 4th International Conference Utrecht, The Netherlands, October 14–17, 2001 Proceedings 4*. Springer. 2001, pp. 343–350 (cit. on p. 23).

[77] J. Kettenbach, G. Kronreif, M. Figl, et al. “Robot-assisted biopsy using ultrasound guidance: initial results from in vitro tests”. In: *European radiology* 15 (2005), pp. 765–771 (cit. on p. 23).

[78] E. C. Pua, M. P. Fronheiser, J. R. Noble, et al. “3-D ultrasound guidance of surgical robotics: A feasibility study”. In: *ieee transactions on ultrasonics, ferroelectrics, and frequency control* 53.11 (2006), pp. 1999–2008 (cit. on p. 23).

[79] V. G. Mallapragada, N. Sarkar, and T. K. Podder. “Robot-assisted real-time tumor manipulation for breast biopsy”. In: *IEEE Transactions on Robotics* 25.2 (2009), pp. 316–324 (cit. on p. 23).

[80] T. R. Nelson, A. Tran, H. Fakourfar, and J. Nebeker. “Positional Calibration of an Ultrasound Image-Guided Robotic Breast Biopsy System”. In: *Journal of Ultrasound in Medicine* 31.3 (2012), pp. 351–359 (cit. on p. 23).

[81] K. Liang, A. J. Rogers, E. D. Light, D. von Allmen, and S. W. Smith. “Three-dimensional ultrasound guidance of autonomous robotic breast biopsy: feasibility study”. In: *Ultrasound in medicine & biology* 36.1 (2010), pp. 173–177 (cit. on p. 23).

[82] C. M. Schneider, A. M. Okamura, and G. Fichtinger. “A robotic system for transrectal needle insertion into the prostate with integrated ultrasound”. In: *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA’04. 2004*. Vol. 1. IEEE. 2004, pp. 365–370 (cit. on p. 23).

[83] L. Phee, J. Yuen, D. Xiao, et al. “Ultrasound guided robotic biopsy of the prostate”. In: *International Journal of Humanoid Robotics* 3.04 (2006), pp. 463–483 (cit. on p. 23).

[84] H. Ho, P. Mohan, E. Lim, et al. “Robotic ultrasound-guided prostate intervention device: system description and results from phantom studies”. In: *The International Journal of Medical Robotics and Computer Assisted Surgery* 5.1 (2009), pp. 51–58 (cit. on p. 23).

[85] W. Ng, V. Chung, S. Vasan, and P. Lim. “Robotic radiation seed implantation for prostatic cancer”. In: *Proceedings of 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Vol. 1. IEEE. 1996, pp. 231–233 (cit. on pp. 23, 24).

[86] Y. Yu, T. Podder, Y. Zhang, et al. “Robotic system for prostate brachytherapy”. In: *Computer Aided Surgery* 12.6 (2007), pp. 366–370 (cit. on p. 23).

[87] R. Kojcev, B. Fuerst, O. Zettinig, et al. “Dual-robot ultrasound-guided needle placement: closing the planning-imaging-action loop”. In: *International journal of computer assisted radiology and surgery* 11 (2016), pp. 1173–1181 (cit. on p. 23).

[88] D. Li, Z. Cheng, G. Chen, et al. “A multimodality imaging-compatible insertion robot with a respiratory motion calibration module designed for ablation of liver tumors: a preclinical study”. In: *International Journal of Hyperthermia* 34.8 (2018), pp. 1194–1201 (cit. on p. 23).

[89] A. I. Chen, M. L. Balter, T. J. Maguire, and M. L. Yarmush. “3D near infrared and ultrasound imaging of peripheral blood vessels for real-time localization and needle guidance”. In: *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part III 19*. Springer. 2016, pp. 388–396 (cit. on p. 23).

[90] D. De Cunha, P. Gravez, C. Leroy, et al. “The MIDSTEP system for ultrasound guided remote telesurgery”. In: *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol. 20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No. 98CH36286)*. Vol. 3. IEEE. 1998, pp. 1266–1269 (cit. on p. 24).

[91] R. P. Budde, R. Meijer, P. F. Bakker, C. Borst, and P. F. Gründeman. “Endoscopic localization and assessment of coronary arteries by 13 MHz epicardial ultrasound”. In: *The Annals of thoracic surgery* 77.5 (2004), pp. 1586–1592 (cit. on p. 24).

[92] C. M. Schneider, G. W. Dachs, C. J. Hasser, M. A. Choti, S. P. DiMaio, and R. H. Taylor. “Robot-assisted laparoscopic ultrasound”. In: *Information Processing in Computer-Assisted Interventions: First International Conference, IPCAI 2010, Geneva, Switzerland, June 23, 2010. Proceedings 1*. Springer. 2010, pp. 67–80 (cit. on p. 24).

[93] C. Schneider, J. Guerrero, C. Nguan, R. Rohling, and S. Salcudean. “Intra-operative “Pick-Up” ultrasound for robot assisted surgery with vessel extraction and registration: a feasibility study”. In: *Information Processing in Computer-Assisted Interventions: Second International Conference, IPCAI 2011, Berlin, Germany, June 22, 2011. Proceedings 2*. Springer. 2011, pp. 122–132 (cit. on p. 24).

[94] S. Billings, N. Deshmukh, H. J. Kang, R. Taylor, and E. M. Boctor. “System for robot-assisted real-time laparoscopic ultrasound elastography”. In: *Medical Imaging 2012: Image-Guided Procedures, Robotic Interventions, and Modeling*. Vol. 8316. SPIE. 2012, pp. 589–596 (cit. on p. 24).

[95] Q. Mei, S. J. Harris, R. D. Hibberd, J. Wickham, and B. L. Davies. “Optimising operation process for computer integrated prostatectomy”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI’99: Second International Conference, Cambridge, UK, September 19-22, 1999. Proceedings 2*. Springer. 1999, pp. 1042–1051 (cit. on p. 24).

[96] P. J. Stolka, D. Henrich, S. H. Tretbar, and P. A. Federspil. “First 3D ultrasound scanning, planning, and execution of CT-free milling interventions with a surgical robot”. In: *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. 2008, pp. 5605–5610 (cit. on p. 24).

[97] S. Heger, T. Vollborn, M. Niggemeyer, T. Mumme, F. Chuembou, and K. Radermacher. “Robot integrated ultrasound geometry-scanning for trackerless bone cement detection in rthr”. In: *2010 IEEE International Ultrasonics Symposium*. IEEE. 2010, pp. 845–848 (cit. on p. 24).

[98] J. Zhan, J. Cartucho, and S. Giannarou. “Autonomous tissue scanning under free-form motion for intraoperative tissue characterisation”. In: *2020 IEEE international conference on robotics and automation (ICRA)*. IEEE. 2020, pp. 11147–11154 (cit. on p. 24).

[99] F. Langsch, S. Virga, J. Esteban, R. Göbl, and N. Navab. “Robotic ultrasound for catheter navigation in endovascular procedures”. In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2019, pp. 5404–5410 (cit. on p. 24).

[100] J. Esteban, W. Simson, S. Requena Witzig, et al. “Robotic ultrasound-guided facet joint insertion”. In: *International journal of computer assisted radiology and surgery* 13 (2018), pp. 895–904 (cit. on pp. 24, 31).

[101] P. M. Torres, P. J. S. Gonçalves, and J. M. M. Martins. “Robotic motion compensation for bone movement, using ultrasound images”. In: *Industrial Robot: An International Journal* 42.5 (2015), pp. 466–474 (cit. on p. 24).

[102] A. Alsinan, M. Vives, V. Patel, and I. Hacihaliloglu. “Spine surface segmentation from ultrasound using multi-feature guided cnn”. In: *CAOS* 3 (2019), pp. 6–10 (cit. on pp. 31, 32).

[103] P. Wang, V. M. Patel, and I. Hacihaliloglu. “Simultaneous segmentation and classification of bone surfaces from ultrasound using a multi-feature guided CNN”. In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part IV 11*. Springer. 2018, pp. 134–142 (cit. on p. 31).

[104] P. Wang, M. Vives, V. M. Patel, and I. Hacihaliloglu. “Robust real-time bone surfaces segmentation from ultrasound using a local phase tensor-guided CNN”. In: *International Journal of Computer Assisted Radiology and Surgery* 15.7 (2020), pp. 1127–1135 (cit. on p. 31).

[105] J. Hetherington, V. Lessoway, V. Gunka, P. Abolmaesumi, and R. Rohling. “SLIDE: automatic spine level identification system using a deep convolutional neural network”. In: *International journal of computer assisted radiology and surgery* 12 (2017), pp. 1189–1198 (cit. on p. 31).

[106] M. Tirindelli, M. Victorova, J. Esteban, et al. “Force-ultrasound fusion: Bringing spine robotic-us to the next “level””. In: *IEEE Robotics and Automation Letters* 5.4 (2020), pp. 5661–5668 (cit. on p. 31).

[107] C. Shorten and T. M. Khoshgoftaar. “A survey on image data augmentation for deep learning”. In: *Journal of big data* 6.1 (2019), pp. 1–48 (cit. on p. 32).

[108] I. Goodfellow, J. Pouget-Abadie, M. Mirza, et al. “Generative adversarial networks”. In: *Communications of the ACM* 63.11 (2020), pp. 139–144 (cit. on p. 32).

[109] A. Zaman, S. H. Park, H. Bang, C.-w. Park, I. Park, and S. Joung. “Generative approach for data augmentation for deep learning-based bone surface segmentation from ultrasound images”. In: *International journal of computer assisted radiology and surgery* 15 (2020), pp. 931–941 (cit. on p. 32).

[110] N. Baka, S. Leenstra, and T. van Walsum. “Ultrasound aided vertebral level localization for lumbar surgery”. In: *IEEE transactions on medical imaging* 36.10 (2017), pp. 2138–2147 (cit. on p. 32).

[111] D. Q. Duong, K.-C. T. Nguyen, N. R. Kaipatur, et al. “Fully automated segmentation of alveolar bone using deep convolutional neural networks from intraoral ultrasound images”. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2019, pp. 6632–6635 (cit. on p. 32).

[112] K.-C. T. Nguyen, D. Q. Duong, F. T. Almeida, et al. “Alveolar bone segmentation in intraoral ultrasonographs with machine learning”. In: *Journal of Dental Research* 99.9 (2020), pp. 1054–1061 (cit. on p. 32).

[113] T. Ungi, H. Greer, K. R. Sunderland, et al. “Automatic spine ultrasound segmentation for scoliosis visualization and measurement”. In: *IEEE Transactions on Biomedical Engineering* 67.11 (2020), pp. 3234–3241 (cit. on p. 32).

[114] B. Hohlmann, J. Glanz, and K. Radermacher. “Segmentation of the distal femur in ultrasound images”. In: *Current Directions in Biomedical Engineering* 6.1 (2020) (cit. on p. 32).

[115] X. Qi, N. Vora, L. Riera, et al. “Automatic Scan Plane Identification from 2D Ultrasound for Pedicle Screw Guidance”. In: *CAOS* 2 (2018), pp. 168–174 (cit. on p. 32).

[116] B. Benjdira, K. Ouni, M. M. Al Rahhal, A. Albakr, A. Al-Habib, and E. Mahrous. “Spinal cord segmentation in ultrasound medical imagery”. In: *Applied Sciences* 10.4 (2020), p. 1370 (cit. on p. 32).

[117] K. Luan, Z. Li, and J. Li. “An efficient end-to-end CNN for segmentation of bone surfaces from ultrasound”. In: *Computerized Medical Imaging and Graphics* 84 (2020), p. 101766 (cit. on p. 32).

[118] H. Patel and I. Hacihaliloglu. “Improved automatic bone segmentation using large-scale simulated ultrasound data to segment real ultrasound bone surface data”. In: *2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*. IEEE. 2020, pp. 288–294 (cit. on p. 32).

[119] C. P. Bridge and J. A. Noble. “Object localisation in fetal ultrasound images using invariant features”. In: *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2015, pp. 156–159 (cit. on p. 36).

[120] O. Ronneberger, P. Fischer, and T. Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer. 2015, pp. 234–241 (cit. on pp. 37, 66).

[121] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. “Densely connected convolutional networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 4700–4708 (cit. on p. 38).

[122] C. Whatmough, J. Guitian, E. Baines, et al. “Ultrasound image compounding: Effect on perceived image quality”. In: *Veterinary Radiology & Ultrasound* 48.2 (2007), pp. 141–145 (cit. on p. 41).

[123] R. N. Rohling. “3d freehand ultrasound: reconstruction and spatial compounding”. dissertation. Churchill College, University of Cambridge, 1998 (cit. on p. 42).

[124] M. H. Mozaffari and W.-S. Lee. “Freehand 3-D Ultrasound Imaging: A Systematic Review”. In: *Ultrasound in Medicine & Biology* 43.10 (2017), pp. 2099–2124 (cit. on p. 42).

[125] V. Grau and J. A. Noble. “Adaptive Multiscale Ultrasound Compounding Using Phase Information”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2005*. Ed. by J. S. Duncan and G. Gerig. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 589–596 (cit. on p. 42).

[126] C. Hennersperger, M. Baust, D. Mateus, and N. Navab. “Computational Sonography”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Ed. by N. Navab, J. Hornegger, W. M. Wells, and A. Frangi. Cham: Springer International Publishing, 2015, pp. 459–466 (cit. on p. 43).

[127] R. Göbl, D. Mateus, C. Hennersperger, M. Baust, and N. Navab. *Redefining Ultrasound Compounding: Computational Sonography*. Nov. 2018 (cit. on p. 43).

[128] A. L. Y. Hung and J. Galeotti. “Good and bad boundaries in ultrasound compounding: preserving anatomic boundaries while suppressing artifacts”. In: *International journal of computer assisted radiology and surgery* 16.11 (2021), pp. 1957–1968 (cit. on p. 43).

[129] C. Berge, A. Kapoor, and N. Navab. “Orientation-Driven Ultrasound Compounding Using Uncertainty Information”. In: vol. 8498. June 2014, pp. 236–245 (cit. on pp. 43, 55).

[130] Q. Huang, J. Lan, and X. Li. “Robotic Arm Based Automatic Ultrasound Scanning for Three-Dimensional Imaging”. In: *IEEE Transactions on Industrial Informatics* 15.2 (2019), pp. 1173–1182 (cit. on p. 43).

[131] C. Yang, M. Jiang, M. Chen, M. Fu, J. Li, and Q. Huang. “Automatic 3-D Imaging and Measurement of Human Spines With a Robotic Ultrasound System”. In: *IEEE Transactions on Instrumentation and Measurement* 70 (2021), pp. 1–13 (cit. on p. 43).

[132] C. Graumann, B. Fuerst, C. Hennersperger, F. Bork, and N. Navab. “Robotic ultrasound trajectory planning for volume of interest coverage”. In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*. 2016, pp. 736–741 (cit. on p. 43).

[133] P. Chatelain, A. Krupa, and N. Navab. “Optimization of ultrasound image quality via visual servoing”. In: *2015 IEEE International Conference on Robotics and Automation (ICRA)*. 2015, pp. 5997–6002 (cit. on p. 44).

[134] V. Sutedjo. “Path Planning Optimization for Robotic Ultrasound Scanning”. Master’s Thesis. Technical University of Munich, 2021 (cit. on p. 45).

[135] A. Karamalis, W. Wein, T. Klein, and N. Navab. “Ultrasound confidence maps using random walks”. In: *Medical image analysis* 16.6 (2012), pp. 1101–1112 (cit. on p. 46).

[136] M. Salehi, S.-A. Ahmadi, R. Prevost, N. Navab, and W. Wein. “Patient-specific 3D ultrasound simulation based on convolutional ray-tracing and appearance optimization”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2015, pp. 510–518 (cit. on p. 49).

[137] J. Vogel, T. Lasser, J. Gardiazabal, and N. Navab. “Trajectory optimization for intra-operative nuclear tomographic imaging”. In: *Medical Image Analysis* 17.7 (2013). Special Issue on the 2012 Conference on Medical Image Computing and Computer Assisted Intervention, pp. 723–731 (cit. on p. 50).

[138] P. Szumowski, M. Mojsak, S. Abdelrazek, et al. “Calculation of therapeutic activity of radioiodine in Graves’ disease by means of Marinelli’s formula, using technetium ((99m)Tc) scintigraphy.” In: *Endocrine* (Dec. 2016) (cit. on p. 60).

[139] J. Chen, H. You, and K. Li. “A review of thyroid gland segmentation and thyroid nodule segmentation methods for medical ultrasound images”. In: *Computer Methods and Programs in Biomedicine* 185 (2020), p. 105329 (cit. on p. 61).

[140] M. Krönke, C. Eilers, D. Dimova, et al. “Tracked 3D ultrasound and deep neural network-based thyroid segmentation reduce interobserver variability in thyroid volumetry”. In: *Plos one* 17.7 (2022), e0268550 (cit. on pp. 62, 66, 71, 93).

[141] A. Guha Roy, S. Conjeti, N. Navab, and C. Wachinger. “QuickNAT: A fully convolutional network for quick and accurate segmentation of neuroanatomy”. In: *NeuroImage* 186 (2019), pp. 713–727 (cit. on pp. 62, 63, 66, 67).

[142] Y. J. Kim, J. H. Seo, H. R. Kim, and K. G. Kim. “Development of a control algorithm for the ultrasound scanning robot (NCCUSR) using ultrasound image and force feedback”. In: *The international journal of medical robotics + computer assisted surgery : MRCAS* 13 (2017) (cit. on p. 62).

[143] Y. R. Stürz, L. M. Affolter, and R. S. Smith. “Parameter identification of the KUKA LBR iiwa robot including constraints on physical feasibility”. In: *IFAC-PapersOnLine* 50.1 (2017), pp. 6863–6868 (cit. on p. 71).

[144] R. Elfring, M. de la Fuente, and K. Radermacher. “Assessment of optical localizer accuracy for computer aided surgery systems”. In: *Computer Aided Surgery* 15.1-3 (2010), pp. 1–12 (cit. on p. 71).

[145] A. M. Franz, T. Haidegger, W. Birkfellner, K. Cleary, T. M. Peters, and L. Maier-Hein. “Electromagnetic tracking in medicine—a review of technology, validation, and applications”. In: *IEEE transactions on medical imaging* 33.8 (2014), pp. 1702–1725 (cit. on p. 71).

[146] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. “OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2019) (cit. on pp. 72, 81).

[147] C. BenMessaoud, H. Kharrazi, and K. F. MacDorman. “Facilitators and Barriers to Adopting Robotic-Assisted Surgery: Contextualizing the Unified Theory of Acceptance and Use of Technology”. In: *PLOS ONE* 6.1 (Jan. 2011), pp. 1–11 (cit. on p. 78).

[148] H. Knoop, J. Raczkowsky, U. Wyslucha, T. Fiegele, G. Eggers, and H. Wörn. “Integration of intraoperative imaging and surgical robotics to increase their acceptance”. In: *International Journal of Computer Assisted Radiology and Surgery* 1.5 (2007), pp. 243–251 (cit. on p. 78).

[149] H. McDermott, N. Choudhury, M. Lewin-Runacres, I. Aemn, and E. Moss. “Gender differences in understanding and acceptance of robot-assisted surgery”. In: *Journal of robotic surgery* 14.1 (2020), pp. 227–232 (cit. on p. 78).

[150] S. D. Sierra Marín, D. Gomez-Vargas, N. Céspedes, et al. “Expectations and Perceptions of Healthcare Professionals for Robot Deployment in Hospital Environments During the COVID-19 Pandemic”. In: *Frontiers in Robotics and AI* 8 (2021) (cit. on p. 78).

[151] P. Vichitkraivin and T. Naenna. “Factors of healthcare robot adoption by medical staff in Thai government hospitals”. In: *Health and Technology* 11.1 (2021), pp. 139–151 (cit. on p. 78).

[152] M. Attia, M. Hossny, S. Nahavandi, M. Dalvand, and H. Asadi. “Towards Trusted Autonomous Surgical Robots”. In: *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 2018, pp. 4083–4088 (cit. on p. 78).

[153] J. Torrent-Sellens, A. I. Jiménez-Zarco, and F. Saigí-Rubió. “Do People Trust in Robot-Assisted Surgery? Evidence from Europe”. In: *International Journal of Environmental Research and Public Health* 18.23 (2021) (cit. on p. 78).

[154] A. K. Hall, U. Backonja, I. Painter, et al. “Acceptance and perceived usefulness of robots to assist with activities of daily living and healthcare tasks”. In: *Assistive Technology* 31.3 (2019). PMID: 29125804, pp. 133–140. eprint: `https://doi.org/10.1080/10400435.2017.1396565` (cit. on p. 78).

[155] E. Broadbent, R. Stafford, and B. MacDonald. “Acceptance of healthcare robots for the older population: Review and future directions”. In: *International journal of social robotics* 1.4 (2009), pp. 319–330 (cit. on p. 78).

[156] Y.-H. Wu, J. Wrobel, M. Cornuet, H. Kerhervé, S. Damnée, and A.-S. Rigaud. “Acceptance of an assistive robot in older adults: a mixed-method study of human–robot interaction over a 1-month period in the Living Lab setting”. In: *Clinical interventions in aging* 9 (2014), p. 801 (cit. on p. 78).

[157] S. Koceski and N. Koceska. “Evaluation of an assistive telepresence robot for elderly healthcare”. In: *Journal of medical systems* 40.5 (2016), pp. 1–7 (cit. on p. 78).

[158] L. Bodenhagen, K. Fischer, and H. M. Weigelin. “The Influence of Transparency and Adaptability on Trust in Human-Robot Medical Interactions”. In: 2017 (cit. on p. 78).

[159] K. Fischer, H. M. Weigelin, and L. Bodenhagen. In: *Paladyn, Journal of Behavioral Robotics* 9.1 (2018), pp. 95–109 (cit. on p. 78).

[160] B. C. Weigelin, M. Mathiesen, C. Nielsen, K. Fischer, and J. Nielsen. “Trust in Medical Human-Robot Interactions based on Kinesthetic guidance”. In: *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 2018, pp. 901–908 (cit. on p. 78).

[161] K. Schaefer. “Measuring Trust in Human Robot Interactions: Development of the “Trust Perception Scale-HRI””. In: Apr. 2016, pp. 191–218 (cit. on pp. 80, 81).

[162] T. Simon, H. Joo, I. Matthews, and Y. Sheikh. “Hand Keypoint Detection in Single Images using Multiview Bootstrapping”. In: *CVPR*. 2017 (cit. on p. 81).

List of Figures

List of Tables