Computer Aided Medical Procedures
Prof. Dr. Nassir Navab

Dissertation

Intelligent Robotic Sonography: Adaptive, Dynamic and Learning-Powered Image Acquisition

Zhongliang Jiang

Fakultät für Informatik
Technische Universität München

Technische Universität München
TUM School of Computation, Information and Technology

Intelligent Robotic Sonography: Adaptive, Dynamic and Learning-Powered Image Acquisition

Zhongliang Jiang

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

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende:	Prof. Dr. Julia Schnabel
Prüfer der Dissertation:	1. Prof. Dr. Nassir Navab
	2. Prof. Dr. Tim Salcudean, Ph.D.

Die Dissertation wurde am 29.03.2022 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 30.08.2022 angenommen.

Zhongliang Jiang
Intelligent Robotic Sonography:
Adaptive, Dynamic and Learning-Powered Image Acquisition
Dissertation, Version 1.1

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

Abstract

Medical ultrasound is crucial for modern clinical intervention. Owing to the advantages of being real-time, affordable, and ionizing radiation-free, ultrasound imaging has been widely used for biometric measurement and diagnosis, in particular of internal organs. However, traditional ultrasound examinations are highly operator-dependent, as it is challenging to accurately maintain or replicate acquisition parameters (the contact force and relative orientation between a probe and a tested object), even for experienced sonographers. The inter and intra-operator variations hamper the diagnosis accuracy; thereby, further limiting the full clinical acceptance of ultrasound imaging.

With the aim to improve the clinical diagnosis accuracy and consistency, an intelligent robotic sonographer system is designed to automatically, accurately, and reliably accomplish image acquisition. Besides the enabling technologies (force control, scanning path planning, and probe orientation optimization) of robotic ultrasound acquisition, this thesis further presents two characterized advanced technologies (motion-aware and deformation-aware) to address the practical challenges of tissue motion and deformation in real scenarios. The motion-aware feature enables the combination of the advantages of free-hand manner (flexibility) and robotic manner (accuracy and stability), which allows the robotic sonography to properly react to the expected/unexpected patient movement during scans. Besides, the deformation-aware characteristic aims to correct the tissue deformation caused by the inevitable pressure applied by the US probe, which facilitates the reproducible and accurate image acquisition of the anatomy of interest. In addition, a novel end-to-end framework is designed for autonomous robotic screening of tubular structures based on real-time ultrasound image feedback. The probe position is controlled to display the object at the image centerline, while the orientation is optimized to be normal to the target vessel during scans. The vascular radius is accurately computed in real-time by solving a constrained non-linear optimization problem, even when the probe is tilted from the normal direction.

Finally, to intuitively transfer high-level physiological knowledge to robotic sonography, an advanced learning framework is designed to enable learning from experts' demonstration. The clinician's intention is inferred by a reward function computed using a novel probabilistic spatial ranking approach, without the requirement of burdensome human annotation. The novel learning-based framework enables autonomous “exploration" of target anatomies based on a few expert demonstrations, without complex modeling and programming. The performance has been validated both on a vascular phantom and more challenging ex-vivo animal organs (chicken heart and lamb kidney).

Zusammenfassung

Medizinischer Ultraschall ist essentiell für moderne medizinische Eintriffe. Dank ihrer Vorteile, nämlich dass sie in Echtzeit verfügbar, erschwinglich und frei von ionisierender Strahlung ist, wird die Ultraschallbildgebung für biometrische Messungen und Diagnostik, im speziellen der inneren Organe, weitgehend eingesetzt. Jedoch sind traditionell durchgeführte Ultraschalluntersuchungen sehr abhängig vom Operateur, da es schwierig ist die Aufnahmeparameter (die Kontaktkraft und relative Orientierung zwischen der Sonde und dem Testobjekt) präzise konstant zu halten oder gewisse Aufnahmeparameter zu replizieren. Dies gilt selbst für erfahrene Sonographen. Diese Schwankungen zwischen und innerhalb der Bediener hindern die diagnostische Genauigkeit; wodurch sie eine weitere klinische Akzeptanz der Ultraschallbildgebung behindern.

Um die Genauigkeit und Konsistenz klinischer Diagnosen zu verbessern wurde ein intelligentes robotisches Sonographiesystem entworfen, welches automatisch, genau und zuverlässig die Bildaufnahme durchführt. Neben den Technologien (force control, Scanpfadplanung und Sondenorientierungoptimierung), welche die Aufnahme von robotischem Ultraschall ermöglichen, präsentiert diese Arbeit auch zwei fortgeschrittene Technologien (motion-aware und deformation-aware) um die praktischen Herausforderungen von Gewebebewegungen und Deformierungen in echten Szenarios zu adressieren. Das motion-aware Feature ermöglicht es die Vorteile der Freihand-Technik (namentlich die Flexibilität) und der robotischen Technik (namentlich die Genauigkeit und Stabilität) zu kombinieren. Dies erlaubt es dem robotischen Sonographer in angemessener Weise auf erwartete oder unerwartete Bewegungen des Patienten während der Aufnahme zu reagieren. Des Weiteren ermöglicht die deformation-aware Methodik Gewebedeformierungen, welche durch den unvermeidbaren Druck der Sonde auf dem Gewebe entstehen, zu korrigieren. Dies erleichtert reproduzierbare und genaue Bildaufnahme der betrachteten Anatomie. Außerdem wurde ein neuartiges Ende-zu-Ende Framework für autonomes robotisches Scannen von tubulären Strukturen basierend auf dem Feedback von Ultraschallbildern in Echtzeit entwickelt. Die Sondenposition wird mittels Regelungstechnik derart geregelt, dass sie das Objekt an der Mittellinie des Bildes anzeigt, während die Orientierung derart optimiert wird, dass sie senkrecht zu der gewünschten Blutbahn während der Aufnahme steht. Der Radius der Blutbahn wird genau in Echtzeit berechnet indem ein beschränktes nicht-lineares Optimisierungsproblem gelöst wird, selbst wenn die Sonde nicht senkrecht zur Anatomie steht.

Zum Schluss wird, um intuitiv hochwertiges physiologisches Wissen auf robotische Sonographer zu übertragen, ein fortgeschrittenes Lernsystem entworfen um das Lernen anhand von Demonstrationen durch Experten zu ermöglichen. Die Intentionen des Arztes werden durch eine Rewardfunktion inferiert indem ein neuartiger probabilistischer spatial ranking Ansatz benutzt wird. Dies geschieht ohne auf aufwändige menschliche Annotationen zurückzugrei-

fen. Das neuartige auf lernen basierende Framework ermöglicht autonome Ërkundung"der Zielanatomie basierend auf nur wenigen Demonstrantionen von Experten ohne eine komplexe Modellierung und Programmierung. Die Leistung wurde sowohl durch ein vaskuläres Phantom, als auch durch anspruchsvollere ex-vivo Tierorgane (Hühnerherzen und Niere eines Lammes) validiert.

Acknowledgments

First of all, I would like to express my sincere gratitude to Prof. Nassir Navab for accepting me to do my research at the chair for Computer Aided Medical Procedures & Augmented Reality (CAMP). Without his supervision, trust, encouragement, and also research freedom over the past four years, this thesis would not have been possible. His vision and close guidance have reshaped my future vision, and the impact has been wide beyond this thesis itself.

I would also like to thank my colleagues Matthias Grimm, Dr. Mingchuan Zhou, Javier Esteban, and Walter Simson. The valuable discussions, encouragements, and suggestions really helped me to be integrated into the team at the beginning of this journey. Thanks, Dr. Thomas Wendler for the valuable support in the purchasing of hardware and creating such a nice research atmosphere in the IFL team. Moreover, I would like to thank Martina Hilla and Sara Erfani for taking care of all the administerial work.

I also appreciate all my coauthors and friends, Yuan Gao, Dr. Marco Esposito, Prof. Walter Stechele, and also the clinical partners, Dr. Med. Markus Kroenke, Dr. Med. Angelos Karlas, and Dr. med. Reza Ghotbi. In addition, a special acknowledgment goes to all students I mentored and cooperated with: Yuan Bi, Felix Duelmer, Yue Zhou, Zhenyu Li, Hanyu Wang, Nehil Danis, Xuesong Li, and Mengxue Shi.

Moreover, I would specifically thank "Manfat Restaurant". I got so much fun and relief there. Thank Jingqiao Lin, Chiufan Tang, Xin Lin, Tina Lin, Annabelle Tang, and all the other friends who met in Manfat. It has become my second family in Munich.

Finally, I would like to thank my parent Chuanhua and Chunyan for their support and patience during this long journey. This also applies to my grandfather Chenyi and deceased grandma Mangxi in 2021. Moreover, my dearest thanks go to my wife Bingrong Zhang, and my daughter Emily Muxi Jiang, for whose continuous support, trust, and joy I am deeply grateful.

Contents

List of Abbreviations

US	Ultrasound	3
RUSS	robotic US system	4
2D	two-dimensional	3
3D	three-dimensional	5
CT	computed tomography	3
MRI	magnetic resonance imaging	3
DOF	degree of freedom	11
VR	virtual reality	11
F/T	force/torque	12
PSO	particle swarm optimization	21
CNN	convolutional neural network	21
TCP	tool center point	25

FFT	Fast Fourier Transform	28
DWT	Discrete Wavelet Transform	26
ROS	Robot Operating System	32
SQP	Sequential Quadratic Programming	40
KF	Kalman Filter	43
PAD	Peripheral Artery Disease	52
ABI	Ankle-Brachial Index	53
CTA	Computed Tomography Angiography	53
ICP	Iterative Closest Point	56
ROI	region of interest	58
IMU	inertial measurement unit	121
RL	Reinforcement Learning	122
IRL	inverse reinforcement learning	122
IL	imitation learning	122

List of Authored and Co-authored Publications

2022

[1] **Zhongliang Jiang***, Yuan Gao*, Le Xie, Nassir Navab. "Towards Autonomous Atlas-Based Ultrasound Acquisitions in Presence of Articulated Motion," *IEEE Robotics and Automation Letters (presented at IROS2022)*, vol. 7, no. 3, pp. 7423-7430, 2022 (directly accepted after the first round of review).

[2] Yuan Bi*, **Zhongliang Jiang***, Yuan Gao, Thomas Wendler, Angelos Karlas, Nassir Navab. "VesNet-RL: Simulation-based Reinforcement Learning for Real-World US Probe Navigation," *IEEE Robotics and Automation Letters (presented at IROS2022)*, vol. 7, no. 3, pp. 6638-6645, 2022 (directly accepted after the first round of review).

[3] **Zhongliang Jiang***, Nehil Danis*, Yuan Bi, Mingchuan Zhou, Markus Kroenke, Thomas Wendler, Nassir Navab. "Precise Repositioning of Robotic Ultrasound: Improving Registration-Based Motion Compensation Using Ultrasound Confidence Optimization," *IEEE Transactions on Instrumentation and Measurement*, Early access, 2022.

[4] Lienhung Chen, **Zhongliang Jiang**, Long Cheng, Alois C Knoll, Mingchuan Zhou. "Deep Reinforcement Learning Based Trajectory Planning Under Uncertain Constraints," *Frontiers in Neurorobotics*, vol. 16, 2022.

2021

[5] **Zhongliang Jiang***, Yue Zhou*, Yuan Bi, Mingchuan Zhou, Thomas Wendler, Nassir Navab. "Deformation-Aware Robotic 3D Ultrasound," *IEEE Robotics and Automation Letters (presented at IROS2021)*, vol. 6, no. 4, pp. 7675-7681, 2021.

[6] **Zhongliang Jiang***, Hanyu Wang*, Zhenyu Li, Matthias Grimm, Mingchuan Zhou, Ulrich Eck, Sandra V Brecht, Tim C Lueth, Thomas Wendler, Nassir Navab. "Motion-Aware Robotic 3D ultrasound," In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 12494-12500, 2021.

[7] **Zhongliang Jiang***, Zhenyu Li*, Matthias Grimm, Mingchuan Zhou, Marco Esposito, Wolfgang Wein, Walter Stechele, Thomas Wendler, Nassir Navab. "Autonomous Robotic Screening of Tubular Structures based only on Real-Time Ultrasound Imaging Feedback," *IEEE Transactions on Industrial Electronics*, Early access, 2021.

[8] Long Lei, Li Huang, Baoliang Zhao, Ying Hu, **Zhongliang Jiang**, Jianwei Zhang, Bing Li. "Diffeomorphic Respiratory Motion Estimation of Thoracoabdominal Organs for Image-Guided Interventions," *Medical Physics*, vol. 48, no. 8, pp. 4166–4176, 2021.

2020

[9] **Zhongliang Jiang**, Matthias Grimm, Mingchuan Zhou, Ying Hu, Javier Esteban, Nassir Navab. "Automatic Force-based Probe Positioning for Precise Robotic Ultrasound Acquisition," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 11, pp. 11200–11211, 2020.

[10] **Zhongliang Jiang**, Long Lei, Yu Sun, Xiaozhi Qi, Ying Hu, Bing Li, Nassir Navab, Jianwei Zhang. "Model-Based Compensation of Moving Tissue for State Recognition in Robotic-Assisted Pedicle Drilling," *IEEE Transactions on Medical Robotics and Bionics*, vol. 2, no. 3, pp. 463–473, 2020.

[11] **Zhongliang Jiang**, Matthias Grimm, Mingchuan Zhou, Javier Esteban, Walter Simson, Guillaume Zahnd, Nassir Navab. "Automatic Normal Positioning of Robotic Ultrasound Probe based only on Confidence Map Optimization and Force Measurement," *IEEE Robotics and Automation Letters (presented at ICRA2020)*, vol. 5, no. 2, pp. 1342–1349, 2020.

[12] Yu Sun, Li Wang, **Zhongliang Jiang**, Bing Li, Ying Hu, Wei Tian. "State Recognition of Decompressive Laminectomy with Multiple Information in Robot-Assisted Surgery," *Artificial Intelligence in Medicine*, vol. 102, pp. 101763, 2020.

[13] Yongfei Yang, **Zhongliang Jiang**, Yuanyuan Yang, Xiaozhi Qi, Ying Hu, Jianjun Du, Bing Han, Guiqin Liu. "Safety Control Method of Robot-Assisted Cataract Surgery with Virtual Fixture and Virtual Force Feedback," *Journal of Intelligent & Robotic Systems*, vol. 97, no. 1, pp. 17–32, 2020.

2019

[14] **Zhongliang Jiang**, Xiaozhi Qi, Yu Sun, Ying Hu, Guillaume Zahnd, Jianwei Zhang. "Cutting Depth Monitoring based on Milling Force for Robot-Assisted Laminectomy," *IEEE Transactions on Automation Science and Engineering*, vol. 17, no. 1, pp. 2–14, 2019.

[15] Yu Sun, **Zhongliang Jiang**, Xiaozhi Qi, Ying Hu, Bing Li, Jianwei Zhang. "Model-based Spinal Deformation Compensation in Robot-Assisted Decompressive Laminectomy," *Mechatronics*, vol. 59, pp. 115–126, 2019.

2018

[16] Yu Sun, **Zhongliang Jiang**, Xiaozhi Qi, Ying Hu, Bing Li, Jianwei Zhang. "Robot-Assisted Decompressive Laminectomy Planning based on 3D Medical Image," *IEEE Access*, vol. 6, pp. 22557–22569, 2018.

2017

[17] **Zhongliang Jiang**, Yu Sun, Shijia Zhao, Ying Hu, Jianwei Zhang. "A Model of Vertebral Motion and Key Point Recognition of Drilling with Force in Robot-Assisted Spinal Surgery," In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 6445–6462, 2017.

---

* represents equal contribution

Part I

Introduction

Introduction

1

> 路漫漫其修远兮，吾将上下而求索
> *Long, long had been my road and far, far was the journey; I would go up and down to seek my heart's desire.*
>
> — **Yuan Qu (c. 340 BC — 278 BC)**
> (Translated by David Hawkes)

1.1 Motivation

Ultrasound (US) imaging was initially inspired by the models of sonar and radar used in World War II [18]. In 1940, supersonic reflectoscope invented by F. Firestone was applied to locate the defects in metals; and then, the commercialized versions of reflectoscope were further used to visualize internal human tissues [18]. Today, US imaging has been one of the most important imaging modalities in clinical practices. In the year from July 2016 to July 2017, over 9.2 million US scans were performed on patients in England, which is almost twice the number of computed tomography (CT) images and three times the number of magnetic resonance imaging (MRI) images in the same period [19]. Compared with CT and MRI, US imaging is radiation-free and real-time, which makes it to be the preferred imaging modality for some clinical applications, e.g., obstetric and pediatric investigations. In addition, the cost of US scans is much less than MRI and CT scans. This characteristic makes it attractive as the first line of investigation in many cases [19].

In traditional free-hand US examination, a US probe is holed by a sonographer and moved manually towards the target US plane for biometric measurement or lesion diagnosis. To properly visualize the target anatomy, sonographers need to actively and continually adjust the probe orientation and applied pressure. This un-ergonomic manner often results in work-related musculoskeletal disorders [20, 21]. In addition, due to the inherited characteristics, US imaging is always suffered from noise and low contrast. It usually requires years to properly train a novice to identify and interpret real-time two-dimensional (2D) images [21]. As a result, the resulting US imaging quality is highly reliant on the sonographer's experience and it could vary a lot between different sonographers or even different examinations carried by the same sonographer at different times [22]. Such a feature severely impairs the consistency and accuracy of the diagnosis results from US images.

To address these challenges of free-hand US examinations, robotic systems have been seen as a promising solution to further achieve high-quality and reproducible US images. Compared with free-hand US examinations, robotic systems can accurately control the probe pose (both position and orientation) and contact force during the scanning. This characteristic enables the feasibility to apply US modality for clinical tasks requiring regular examinations, e.g., monitoring the tumor size [23]. Besides, the use of robotic manipulators can help address some tricky socialized issues, like the unbalanced distribution of health care resources between developed and developing areas. Furthermore, robotic US system (RUSS) can help reduce the waiting time for patients. This feature is very important for identifying the disease in the mild (early) phase [24, 25].

The RUSS can be categorized into teleoperation and autonomous groups. The teleoperation approach has been seen as one of the most mature and successful techniques applied to medical robotic systems. The most successful representative is the da Vinci surgery system (Intuitive Surgical Inc.), which has been approved to be used for clinical treatment of cardiac, colorectal, general, gynecology, urology, etc [26]. The slave robotic manipulator at the patient's site is fully controlled by the clinician at the remote control console and the responsibility of the surgery is fully held by the operators [27]. This manner is the key to resulting in regulatory approval. Specific to teleoperation RUSS, a few examples have been successfully deployed at various hospitals or rural areas for remote US examinations [28, 29]. Nevertheless, some key acquisition parameters (e.g., probe orientation and scanning trajectory) are still manipulated by human operators. This factor hinders the achievement of repeatable US images.

In order to further enable robust and reproducible US scans, autonomous RUSS has attracted increased attention in the past two decades. The deployment of such autonomous systems can alleviate the requirement of experienced sonographers, who are usually rare in undeveloped and rural areas. The development of intelligent RUSS requires a high level of interdisciplinary knowledge related to the fields of robotic control, computer vision, US imaging processing, and artificial intelligence. Although autonomous RUSS has great potential, there are no commercial products in the market yet. To bridge the gap between engineering developments and clinical needs, more practical factors need to be taken into consideration. This dissertation explores the way to automatically optimize the probe orientation, address the potential object movement and unavoidable force-induced imaging deformation, and enable the ability to learn how to perform US examinations.

1.2 Outline

This dissertation consists of three parts. In **Part I**, **chapter 1** the basic concepts of US imaging and an overview of robotic US imaging with a focus on the development of autonomous systems, are introduced.

In **Part II**, the contributions of this dissertation are presented in details:

- **Chapter 2** describes the methods aiming at **optimizing the probe orientation**, i.e., aligning the US probe centerline to the object's surface normal at the point of contact in order to improve sound propagation within the underlying tissues. The ability to automatically identify the normal direction of contact surface without requirements of prior knowledge is important for reproducible US scans and parameterizing the clinical knowledge of experienced sonographers.

- **Chapter 3** proposes a **motion-aware RUSS** aiming at achieving complete and accurate three-dimensional (3D) US volumes of non-stationary imaged objects during US sweeps. This characteristic is very important in clinical practices. The potential object movements during the scanning will decrease the quality of 3D compounding results, particularly for the anatomies with long structures. Furthermore, to properly visualize the whole anatomy, e.g., displaying the entire limb artery tree, attempted adjustments of the object are necessary during the scanning. To be aware to the object movement and further compensate for this motion to continue the scanning process, a depth camera is employed to track the position of the firmly attached passive markers in real-time.

- **Chapter 4** describes a **deformation-aware RUSS** aiming at recovering zero-compression 3D geometry of the anatomies from the deformed images. This study facilitates the accurate biometric measurement and diagnosis using US modality. To achieve this objective, image pixel displacements, nonlinear tissue stiffness, and contact forces are used to build an empirically polynomial regression. The pixel-wise displacement is obtained using an optical flow algorithm. The patient-specific stiffness is estimated by performing robotic palpation. Due to the use of tissue stiffness, the optimized regression model can be applied at unseen points or even unseen objects by substituting the estimated local stiffness. The corrections are carried on individual 2D images; and then finally reconstructed the zero-compression 3D geometry using the corrected 2D images.

- **Chapter 5** demonstrates an end-to-end workflow for **autonomous robotic screening of tubular structures** based only on real-time US images. The manner of autonomous acquisitions benefits for eliminating the inter-operator variability; thus, resulting in accurate diagnosis and staging of vascular diseases. The US probe is required to be roughly placed on the target tissue. Then, the proposed RUSS can automatically screen along the targeted blood vessel with an optimized pose and estimate the radius of the vessel in real-time. To achieve this objective, a U-Net structure is employed to segment the blood vessel from cross-sectional US images. Then, the 3D boundary of a local section of the target tubular tissue is depicted using the segmented binary masks of the most recent several B-mode images. The centerline of the target structure and its local radius can be computed by solving a constrained non-linear optimization problem based on the extracted local 3D point clouds.

- **Chapter 6** describes a **leaning-based RUSS** with the ability to autonomously “explore” the standard scan planes defined by clinicians through demonstrations. The demonstrations starting from random positions and ending at the target objects are manually given by experts. In this manner, the high-level physiological knowledge from human sonographers can be intuitively transferred to the robotic sonographer. To understand the underlying intention of human sonographeres, a reward function for assessing individual US images is learned from demonstrations. To avoid inconsistent objectives, all demons are ranked and cleaned using Gaussian distribution. The inferred rewards of individual B-mode images are then used to locate target anatomies by successively performing coarse and fine localization. To validate the performance of the proposed method, experiments have been carried out in a simulated grid world and on challenging physical phantoms using ex-vivo animal organ phantoms (lamb kidney and chicken heart).

In each of these chapters, the performance of the presented methodology is not only validate using gel phantoms/simulation environments but also evaluated on ex-vivo animal organs (Chapter 6) and in-vivo human volunteers (Chapter 2, 4, 5). Substantial parts of this dissertation have already been published. The related publications and my personnel contributions are clearly summarized at the beginning of each chapter. Although the main ideas and the detailed implementations of this work were done on my own if not explicitly declared otherwise, all this work will be not possible without the great contributions and insightful discussions with fellow colleagues and coauthors.

Finally, **Part III**, **Chapter 7**, summarizes the findings of this dissertation, and a discussion about the open challenges and the future perspectives of autonomous RUSS are discussed in **Chapter 8**.

1.3 Ultrasound Imaging Physics

1.3.1 Transducers

The transducer (US probe) is the device that generates US signals from electrical transmission pulses and converts the received US signals into electrical echo signals [19]. The linear and curvilinear transducers are the two most common types used for general applications in clinical practices. Besides these two types of probes, the phased-array probe, 3D probe, endo-cavity probe are developed for different applications. Although only the linear and curvilinear probes were used in this dissertation, we summarized various types of probes in this Section.

Linear and Curvilinear Transducers

The demonstration of linear and curvilinear scan formats is depicted in Fig. 1.1 (a)-(b). On the one hand, a linear probe has a rectangular imaging view and a linear transducer face, which make it particularly suitable for the anatomy close to the surface, e.g., neck and limbs. On the other hand, the electrodes are deployed along the front face in a shape of an arc, instead of a straight line. This characteristic leads to the advantage that the field of view becomes wider with depth. Thereby, the curvilinear is often used for abdominal applications requiring deep view.

Phased Transducers

Besides the linear/curvilinear probes, phased-array probes [see Fig. 1.1 (c)] with a small footprint is developed to visualize a larger underlying area. The phased probes are most commonly used in cardiac imaging owing to the advantage that displaying a large area without moving probe pose. Compared with linear and curvilinear probes, the individual elements are much narrower, which allows all elements in the probe tip face are activated for every scan line.

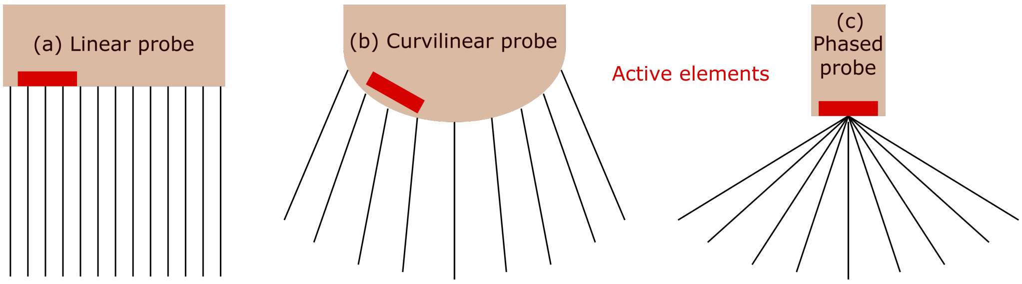


Fig. 1.1. Illustration of a linear, curvilinear and phased scan mode. The red block represents the current active elements.

Thereby, a fast beam shifting performance benefits for the tracking of the respiratory-induced tissue motions e.g., liver and kidney [30].

Endo-Cavity Transducers

The aforementioned probes are used outside the human body. To further improve the imaging quality of some internal anatomies, e.g., prostate and cardiac tissues, endo-cavity probes with round shapes are designed. These probes can be inserted into a natural body cavity or a surgical opening to position the probe close to the anatomies of interest. Due to the characteristics of US modality, high-quality imaging can be achieved because the attenuation effects caused by intervening tissues can be reduced. The four representative endo-cavity probes have been depicted in Fig. 1.2 [19].

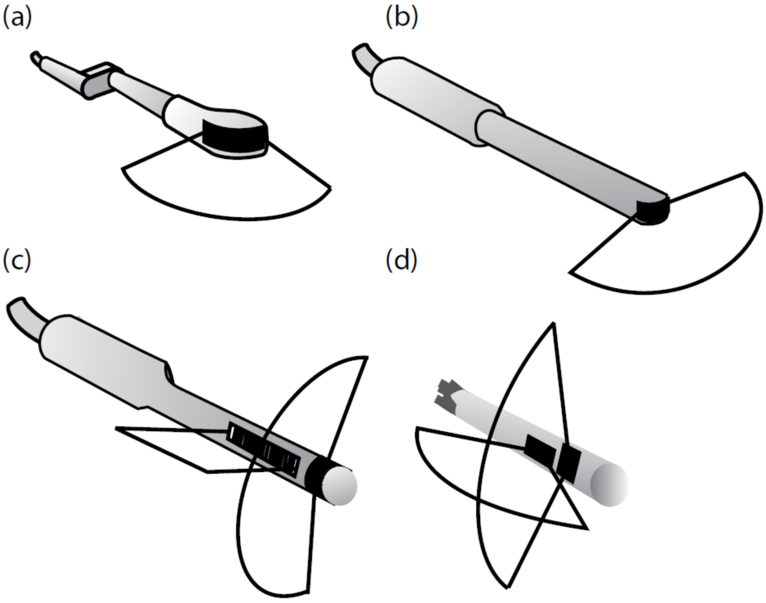
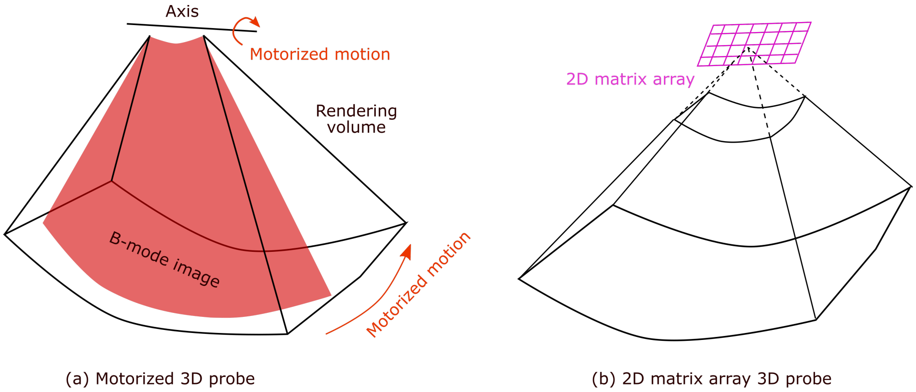


Fig. 1.2. Examples of four representative endo-cavity transducers [19]. (a) Curvilinear transducer for transvaginal scanning; (b) curvilinear transducer (small face area) for trans-rectal or trans-vaginal scanning; (c) “Bi-plane" trans-rectal transducer allowing both transverse and longitudinal scans of the prostate; (d) Trans-oesophageal transducer with two phased arrays giving two orthogonal cross-sections of the heart. *Reprinted from [19] with kind permission from Taylor & Francis Group books.*

3D Transducers

In order to display the object in 3D for intuitive visualization, various 3D probes are developed. There are two common ways: 1) motorized 3D probe and 2) matrix array 3D probe. The illustration of these two types probes is shown in Fig. 1.3. Regarding the motorized version, elements are deployed in a linear array, thereby, 2D images can be acquired in real-time like the linear/curvilinear probe. To build 3D volume, a motor is often employed to rotate the 2D array fully imaging a pyramidal volume of tissues. However, due to the mechanical nature of the motorized movement and line-by-line scanning of each slice, the acquisition rate of motorized 3D probes is limited. Regarding matrix 3D probes [see Fig. 1.3 (b)], thousands of elements are distributed in a 2D matrix. Then, the 2D sector can be swept to visualize a pyramidal volume and similar resolutions can be achieved in both lateral (field of the US view) and elevation direction [31].



(a) Motorized 3D probe

(b) 2D matrix array 3D probe

Fig. 1.3. (a) motorized 3D probe, and (b) 2D matrix array 3D probe.

1.3.2 Ultrasound Waves Propagation

In this Section, the basic principles of US imaging are briefly described.

B-mode Image

B-mode (brightness mode) is one of the most common US imaging modalities, which provides a cross-sectional view of underlying tissues and organ boundaries (see Fig. 1.4). The brightness value at each pixel is related to the strength of the reflected echo. After the US elements “fire" the US signal, the transducer will receive a range of echoes, generated by reflection and scatteration of US wave at object boundary. Each echo is displayed at the position corresponding to the interface or tissues that produced it. These relative position in images d is estimated using the pulse-echo principle as Eq. (1.1).

$$d = \frac{ct}{2} \tag{1.1}$$

where c is the speed of sound, and t is the total time that echo arrives back at the transducer. c is often set to $1540\ m/s$ for human tissues.

Regarding non-homogeneous human tissues, there are numerous interfaces or irregularities. Thus, multiple echoes will be generated by these interfaces and scatter when the emitted

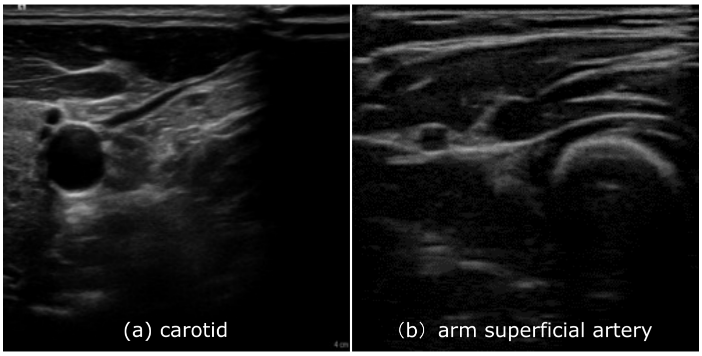


Fig. 1.4. (a) a B-mode image of a volunteer's carotid acquired using a Siemens Acuson Juniper US machine, and (b) a B-mode image of a volunteer's forearm acquired using a Cephasonics US machine.

pulse travels through of human body. The sequence of echoes arrives transducer orderly in terms of the distance d, which forms a line in the B-mode images. The brightness of the line at individual depths is determined by the strength of the echo. To further formalize a 2D image, a large number of B-mode lines are required to guarantee the lateral resolution. An intuitive formation of 2D B-mode image is depicted in Fig. 1.5. The beam is moved step-by-step from one side to the other side of the US imaging view. The number of the elements deployed on the probe face can be determined by balancing the lateral resolution and the time efficiency for a complete sweep of all US elements.

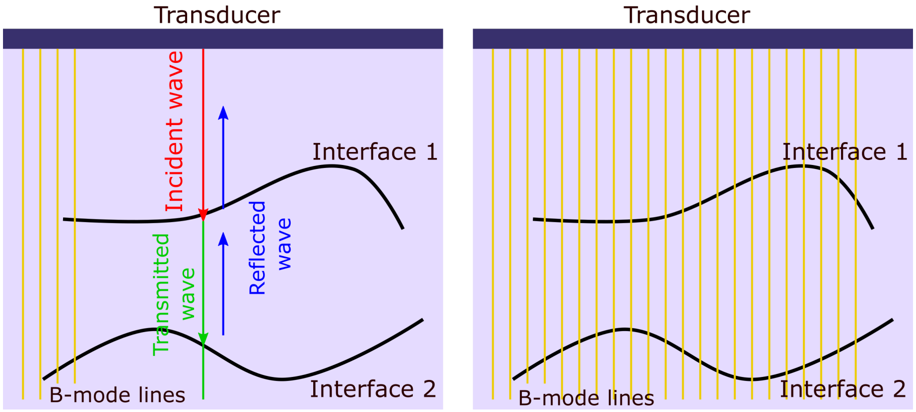


Fig. 1.5. Illustration of B-mode imaging formation based on the line-by-line manner.

Acoustic Impedance

The sound waves with great frequency ($> 20\ kHz$) are referred to as US waves, which usually cannot be heard by human. The US frequencies used most commonly in diagnosis are between $2 - 15\ MHz$ [19]. Thereby, the wavelength λ can be calculated using the sound speed c and frequency f as Eq. (1.2).

$$\lambda = \frac{c}{f} \tag{1.2}$$

Based on Eq. (1.2), the wavelengths in human tissue varied in $[0.1,\ 1\ mm]$ when the frequency set in the range of $2-15\ MHz$. A short λ is helpful to improve the lateral resolution of resulting B-mode images.

Due to the natural characteristics of waves, the incident waves will be partially reflected at tissues interface where is a change in acoustic properties of tissues as Fig. 1.5. Only part of the US signal can pass though the interface into deep tissue, while other energy is reflected back to the source of wave. To investigate the ratio between reflected energy and transmitted energy, the important concept acoustic impedance z is introduced. Acoustic impedance is similar to electrical impedance, which represents the ratio of the voltage and the resulting electrical current. For different materials, the acoustic impedance z can be expressed as Eq. (1.3).

$$z = \rho c \tag{1.3}$$

where ρ represents the material density. The typical z of human tissues or organs are: liver 1.66×10^6; kidney 1.64×10^6; fat 1.33×10^6, water 1.48×10^6, bone 6.47×10^6 and air 430 [19]. The unit is $kgm^{-2}s^{-1}$.

Ultrasound Reflection and Transmission

The amplitudes of the transmitted and reflected waves are determined based on the z of the two tissues around the interface. If the change of z between two materials is very large, the waves will be strongly reflected back. Conversely, most of the wave will be transmitted into the second medium if the change of z between two materials is very small. Then the relationship between reflected pressure p_r and incident pressure p_i can be expressed as Eq. (1.4).

$$\frac{p_r}{p_i} = \frac{z_2 - z_1}{z_2 + z_1} \tag{1.4}$$

where z_1 and z_2 are the acoustic impedance of the two tissues located at the upper and downsides of the interface. The ratio of p_r to p_i is also called amplitude reflection coefficient R_A. The liver-kidney is 0.006, liver-bone is 0.59 and liver-air is 0.9995 [19].

The large R_A for liver-air interface explains the reason of US gel in clinical practices. The use of gel can avoid the air between the probe and contact skin surface, thereby, the US signal can be successfully transmitted into human tissues rather than be reflected away. Similarly, due to the large R_A between soft tissue and bone, US imaging is not able to effectively visualize the anatomies rightly below the bone. Besides the amplitude of reflection coefficient R_A, reflection coefficient can also be described in terms of intensity reflection coefficient R_I, which can be calculated as Eq. (1.5).

$$R_I = R_A^2 = \left(\frac{z_2 - z_1}{z_2 + z_1}\right)^2 \tag{1.5}$$

1.4 Robotic Ultrasound Imaging Systems

Robotic US imaging is attracting increasing attention recently. The requirement for RUSS has received a boost by the current COVID-19 pandemic [32, 33]. The use of robotic techniques in hospitals allows separating the patients and sonographers. This is extremely beneficial for reducing risks of virus transmission for patients and medical staff. To emphasize the contributions of this dissertation, the development history of existing researches has been described in this Section.

1.4.1 Teleoperation RUSS

Teleoperation techniques enable operators to remotely perform certain tasks. Due to the development of networks, multimedia, and communication technologies, teleoperation becomes one of the most mature techniques in the field of medical robotics [33]. Regarding a tele-echosonography system, a robotic manipulator is often employed to hold a US probe on the slave side, while a sonographer at the master side can remotly control the probe for US examinations.

In 2001, Masuda *et al.* proposed a 3D mechanism to hold a US probe while the three orthogonal rotations and translations were remotely controlled by two joysticks [34]. To ensure patient's safety, four sensors were symmetrically deployed to monitor the real-time contact force. However, the missing of force feedback to operators make the operation feeling is different from the sonographer's habit. To address this challenging, Mitsuishi *et al.* proposed a custom-designed six degree of freedom (DOF) mechanism as the master manipulator, which can reflect the remote contact force between probe and the contact surface to operators [35], and this system has been validated remotely at a distance of $700\ km$.

To improve the stability and accuracy of remote RUSS, an image servoing approach was developed to control the three DOF in the field of US view, while the remaining DOFs were still controlled by human operators [36]. The imaging servoing method can provide a pixel-wise control accuracy, thereby, it can restrain the negative influence of human tremor. Considering the patients' safety, Vilchis *et al.* designed a cable-driven nonrigid robot [37]. This system has been used on 100 patients with abdominal aortic aneurysms (AAA) at a distance of $1125\ km$. Alternatively, Tsumura *et al.* proposed a passive mechanism based on springs for fetal examinations, which results in a hard limitations of applied force [38]. Furthermore, regarding the potential time delay caused by limited communication speed between the master and slave sites, a continuous-path controller was developed to realize smooth motion of the slave robotic arm [39].

Besides aforementioned tele-echography systems, Ito *et al.* proposed a portable and attachable robotic system, which has the advantage to be easily handled by paramedics [40]. This feature is very important for emergency medical treatment. To provide intuitive operation feeling, a virtual reality (VR) based simulator was designed as a new type of interface for tele-echography system [41]. The initial evaluation of such a system has been performed by

12 experienced sonographers and the results suggest that the novel interface is suitable for tele-echography system.

1.4.2 Autonomous RUSS

Tele-echography systems have achieved great success; however, some key acquisition parameters (e.g., contact force, probe orientation and scan path) are still manipulated by human operators. To further achieve repeatable and operator-independent US images, various RUSS towards autonomous scanning have been proposed. Due to the inherited characteristic of US modality, a certain force is required to be applied between the US probe and the tested objects. This force vary between $1-20\ N$ depending on the examined tissues [32]. In order to guarantee the patient's safety and to avoid inconsistent deformation, the applied force is desired to be constant during scanning. To achieve this, Gilbertson *et al.* implemented a hybrid position/force controller for a 1-DOF hand-held RUSS [42]. A similar hybrid position/force method using an external force/torque (F/T) sensor was proposed for 6-DOF RUSS [43, 44].

Besides, a decoupled external hybrid force/position controller was employed for a 7-DOF RUSS [45]. Since this approach doesn't require mode changing, the risk of jerking and instability has been maximally limited. Furthermore, implementation of such an external force controller is deemed simpler and can be easily adapted to any kind of robot [45]. Similar as [45], Ma *et al.* used a PID controller to compute the variation of Cartesian position based on the force feedback; and then the desired movement was performed using a position controller provided by the manufacture [46]. To further improve the safety of patients, compliant control methods are employed for the development of RUSS [5, 7, 9, 47, 48, 49, 50, 51, 52, 53], which enables "soft" interactions between patients and RUSS. In addition, Tsumura *et al.* proposed a spring-based mechanism to maintain the contact force and passively adjust the probe pose with respect to the contact surface specifically for fetal examinations [38]. Similarly, Wang *et al.* proposed a customized clutch joint mechanism to physically limit the maximum contact force applied on subjects [54]. However, such passive systems cannot precisely control the contact force.

Besides the contact force, the relative probe orientation with respect to the contact surface is also a key factor dominating the resulting imaging quality. In general cases, a US probe is positioned perpendicularly to the contact surface [55]. This can be explained by the fact that more echoes are reflected back to the US transducer rather than scattered away. To compute the probe pose for scans, depth camera-based approaches are most often used in the existing researches [6, 46, 56, 57, 58]. Based on the object surface point cloud extracted by an R-GBD camera, the probe centerline was aligned to the computed normal direction of the contact surface. The advantage of these approaches is high computation efficiency, while the main limitation is the low accuracy of the estimation results, particularly considering the probe hasn't been positioned on the examined object. Besides the camera-based approaches, Chatelain *et al.* employed visual servo technique to optimize the probe orientation using confidence map [59, 60]. The confidence map provides a pixel-wise measure of signal loss based on a simplified model of sound propagation in examined tissues [61]. Then, Chatelain *et al.* further extended their approach from 2D probe to a 3D wobbler probe [62]. In addition, Huang *et al.* optimized the in-plane orientation based by balancing the contact forces applied at two endpoints on the probe tip using two pressure sensors [56].

To automatically accomplish robotic scans, the scan trajectory for robotic arm is important, particularly when there are obstacles between probe and target anatomies, i.e., ribs. To properly locate and evaluate a artery, Merouche *et al.* maually give the scanning path in the robotic base frame [63]. Besides, Akbari *et al.* proposed a fully US-based approach for breast US acquisition [64]. To achieve this, a pre-scan was carried first manually; and then the post-scan was done to refine the initial trajectory recorded during pre-scan. However, this approach requires a experienced sonographer to carefully determine a proper scan path in real scenarios. To address this issue, different preoperative images (e.g., MRI, CT, and camera images) were used to compute suitable scan paths. Hennersperger *et al.* planned the scan path for abdominal applications by manually selecting the start- and endpoint from MRI data [48]. Based on the registration between the MRI data and the 3D RGB-D surface image acquired using a Kinect camera (Microsoft Corporation, USA), RUSS can automatically obtain the execution path for robotic controller. Similarly, Langsch *et al.* extracted the tracking trajectory of an aorta from a pre-operative MRI to automatically visualize a catheter [65].

There are some survey articles on the state-of-the-art RUSS that have been accomplished to summarize the related literature over the years. The interested reader can refer to[66, 67, 68, 69, 70]. Recently, Von Haxthausen *et al.* further summarized the most recent publications $(2016-2020)$ in the field of RUSS [71], while Li *et al.* purely focus on the development of autonomous system [72].

Part II

Methodology and Contributions

Robotized US Probe Orientation Optimization

2

2.1 Overview and Publications

This chapter presents my contributions towards autonomous normal positioning of robotic US probe. The probe orientation is one of the dominating parameters governing the quality of US images. To develop automatic RUSS, it is important to propose the approaches that allow automatically positioning of the probe orientation with respect to the contact objects. The ability to identify the relative orientation between the probe and the tested object can further result in repeatable and automatic acquisitions. The presented methods aim at aligning the central axis of the US probe to the tissue's surface normal at the point of contact in order to improve sound propagation within the tissue. Thereby, a US image with better contrast is achieved.

After outlining the potential clinical needs of orthopedic applications (Sec. 2.2), this chapter presents the developed methodologies for autonomous normal orienting of US probe in Secs. 2.4 and 2.5. To identify the normal direction of the contact surface without any prior knowledge, the optimization processes are accomplished by carrying out two fan rotations around two orthogonal axes, namely in-plane orientation and out-of-plane orientation. Due to the accessibility of an accurate external force sensor, two distinct approaches have been developed for compute the normal direction of the tested surface:

1. The **image and force based method** as explained in Sec. 2.4 computes the optimal in-plane orientation using the US confidence map [61], and the out-of-plane orientation using the external forces estimated from joint torques. Due to the estimated force is not accurate during the fan rotation, a new feature – smooth derivative of contact force (SDCF) – is defined and used to robustly identify the normal direction of the tested objects at the contact point.

2. The **mechanical model based method** as explained in Sec. 2.5 optimizes the in-plane and out-of-plane orientation purely based on the measured force from an external accurate F/T sensors. To this end, the mechanical model is first built to describe the reaction force during two rotations around a given contact point. Then, the parameters of the derived model are optimized using particle swarm optimization (PSO) [73] based on the measured forces. Due to the derivation of the accurate mechanical model of reaction force, the identification of object's normal direction can be improved.

Substantial parts of this chapter have already been published and quoted verbatim. The methodology, implementation details, and results of the **image and force based approach** for the case that there is no external force sensor involved have been published in the following article.

[11] **Zhongliang Jiang**, Matthias Grimm, Mingchuan Zhou, Javier Esteban, Walter Simson, Guillaume Zahnd, Nassir Navab. "Automatic Normal Positioning of Robotic Ultrasound Probe based only on Confidence Map Optimization and Force Measurement," *IEEE Robotics and Automation Letters (presented at ICRA2020)*, vol. 5, no. 2, pp. 1342–1349, 2020.

Copyright Statement. ©2020 IEEE. With kind permission of IEEE.

The related contents of the **mechanical model based method** have been published in the following article.

[9] **Zhongliang Jiang**, Matthias Grimm, Mingchuan Zhou, Ying Hu, Javier Esteban, Nassir Navab. "Automatic Force-based Probe Positioning for Precise Robotic Ultrasound Acquisition," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 11, pp. 11200–11211, 2020.

Copyright Statement. ©2020 IEEE. With kind permission of IEEE.

2.2 Introduction

2.2.1 Clinical Motivation

The importance of orthopedic surgery amplifies with an increase in life expectancy, as elderly people are more prone to bone-related injuries [74]. For example, 18% of elderly people and 25% of children in the USA suffer from distal radius fracture [74, 75]. To optimize the outcome of an orthopedic surgery, surgical navigation has been widely implemented in current clinical practice. To enable accurate navigation, it is necessary to carry out a registration providing the consistency between the intra-operative US scans and the high resolution pre-operative scans, such as MRI or CT [76, 77]. Regarding orthopedic surgery, the bone surface is often used to perform such a registration, as it has strong resistance to deformation [77]. Therefore, it is crucial to obtain high-quality US bone imaging. However, US examinations are highly operator-dependent, as it is challenging to exactly maintain or replicate acquisition parameters (the contact force and relative orientation between a probe and a tested object), even for experienced sonographers [42].

The relative tilt angle θ_t form the local normal direction is crucial to properly visualize the bone surface in US imaging view. After investigating the influence of θ_t, Hnatsenka *et al.* found that better quality (contrast) could be obtained when θ_t is close to zero (the probe centerline is orientated closely to the normal direction of the contact surface) [55]. This phenomenon is further aggravated in orthopedic applications, as the intensity reflection coefficient for most interfaces from soft tissues to bones are approximately 25% while that for most interfaces from soft tissues to soft tissues are less than 0.01% [19]. Thereby, it is reasonable to place the probe along the normal direction of unknown constraints for general US applications.

To intuitively demonstrate the impact of the US probe orientation on the resulting B-mode images, the images of a volunteer's radius bone in various tilt angles has been illustrated in Fig. 2.1. It can be seen from Fig. 2.1, the radius bone boundary gradually disappears in the US images when the US probe is rotated away from the normal direction of the contact surface. The bone boundary becomes very weak in the case that the tilted angle is over $15°$ as shown in Figs. 2.1 (d), (e), and (f). Therefore, the probe orientation is deemed important to obtain high-quality US bone imaging for the potential registration between intra-operative US imaging and pre-operative scans (CT or MRI).

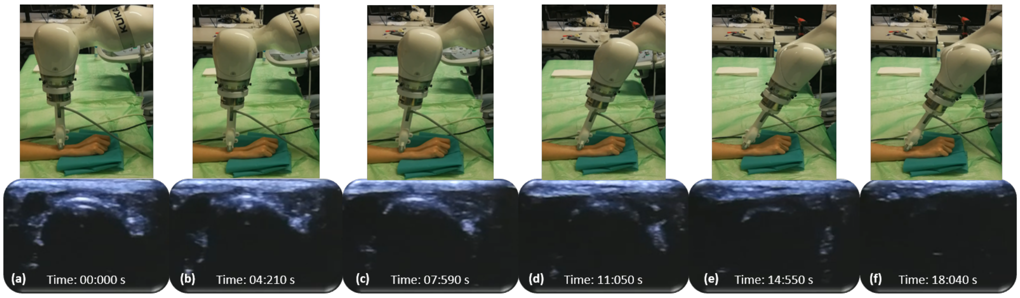


Fig. 2.1. The impact of the US probe orientation on the US scans of the radius bone. The upper row depicts the probe orientation. The lower row shows the corresponding US images of a volunteer's radius bone. (a), (b), (c), (d), (e), and (f) are the images acquired when the linear probe is titled around $0, 5, 10, 15, 20, 25°$, respectively. *Reprint from [9] with permission of IEEE.*

2.2.2 Adjustment of Robotic US Probe Orientation

In the field of industry, the task like robotic polishing also requires to orient a tool attached to the flange along the normal direction of the constraint surface or a given direction θ_t. However, the constrained surface is usually known in industrial applications, such as mold manufacturing. Therefore, the desired posture of attached tools can be directly generated from CAD/CAM software for polishing a freeform surface [78, 79, 80]. Regarding an unknown constraint surface, Merlet *et al.* and Kazanzides *et al.* employed reaction force to identify the surface normal [81, 82]. However the estimation accuracy cannot be guaranteed when the surface is non-rigid. Besides, Yoshikawa *et al.* suggested moving the tool on the surface using a hybrid position/force control, and then, the normal direction was approximated according to the direction that is perpendicular to the line connecting two nearby positions on a considered trajectory [83]. However, this method requires a change of the contact point on the constraint surface. Furthermore, the human objects are soft, a certain deformation will be generated during the scans. This makes the aforementioned approaches originally developed for industrial application cannot be directly transferred for the robotic US application.

Regarding the robotic US applications, Chatelain *et al.* used a visual servoing technique to position a 3D US probe in the normal direction of an unknown constraint surface [62].

However, for 2D probes which are much more common in clinical practice, the method is only able to optimize the in-plane orientation while the out-of-plane part still required to be adjusted via tele-manipulation [59]. Huang *et al.* approximated the normal direction of the constraint surface considering the normal direction of a triangle composed by three neighbor points around the planned path on an image obtained by a depth camera prior to US scanning [56]. But the surface obtained before the probe makes contact with the tissue can not adapt to deformations induced by the contact. Furthermore, the accuracy of the method is limited by the employed hand-eye calibration and potential occlusions between the camera and the scanned tissue.

Without the ability to properly position the probe orientation, RUSS cannot automatically obtain high-quality scans from non-flat surfaces such as human limbs. In order to develop an automatic RUSS providing good image quality, the capability to ascertain the normal direction between probe and surface without requiring external help from human operators is crucial. Since it is independence to calibration accuracy and it is sensitivity to changes in contact condition [14], the contact force F_c is considered as a promising signal for identifying the normal direction $\mathbf{N}$ of contact surface.

2.2.3 Contributions

Considering the two cases that whether an accurate external F/T sensor is involved, two distinct approaches (the **image and force based approach** and the **mechanical model based approach**) are developed. Due to the use of an external F/T sensor, the mechanical model based approach achieved more accurate estimation of the normal direction of contact surface. However, for the applications without requirement of super high accurate identification results ($< 5^\circ$), it will be more economically efficient using the **image and force based approach**. Accurately positioning the US probe with respect to the contact surface is a key characteristic towards automatic RUSS for high-quality US imaging, which can provide accurate and repeatable probe orientation making scans easily comparable. Thereby, US modality can be extended to monitor the changes of the anatomies and lesions of interest. The implementation procedures for two distinct approaches have been summarized as follows:

1. The **image and force based method**: In order to identify the normal direction $\mathbf{N}$ of the contact surface for any given point of contact, the estimated forces and the US images are used. The use of only force measurements and US images is inspired by the way US sonographers conduct US examinations. They do not look at the patient but only at the US images. To this end, the in-plane component $\mathbf{N}_i$ based on the confidence map of the real-time US image is first computed. Then the out-of-plane component $\mathbf{N}_o$ is computed by performing a fan motion around the out-of-plane direction. The estimated external force from joint torques is measured in real-time and analyzed offline.

2. The **mechanical model based method**: In order to further improve the identification accuracy and provide a generic approaches working for both linear and convex probe, a model based orientation optimization method to position the probe perpendicular to the object surface is proposed. To this end, a mechanical model is proposed to describe the reaction force during two rotations around a given contact point. Then,

the unknown parameters of the derived model are optimized using particle swarm optimization (PSO) on the force data measured from an external F/T sensor. The desired normal is decomposed into two parts: an in-plane component $\mathbf{N}_i$ (co-planar with the US image plane) and an out-of-plane component $\mathbf{N}_o$ (orthogonal to the US image plane). These two components are estimated separately by executing two rotations around the corresponding axes, as shown in Fig. 2.2 (b).

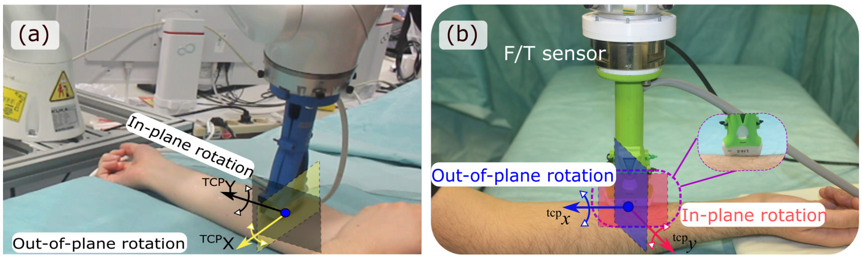


Fig. 2.2. Two rotations executed to optimize the probe orientation when (a) there is no F/T sensor, and (b) there is one. *Reprint from [11] and [9] with permission of IEEE.*

For both novel autonomous normal positioning approaches, results of quantitative validations on gel-wax phantoms and volunteer's tissues, i.e., forearm, upper arm, and lower back, demonstrated that the proposed approaches can accurately estimate the normal direction of the contact surface without any requirement of prior knowledge.

2.3 Effect of Probe Orientation on Bone Detection

2.3.1 Ultrasound Bone Detection

To quantitatively assess the effect of the probe orientation on the quality of US imaging of orthopedic object, the bone surface detection method proposed by Salehi *et al.* [77] is employed in this work. The method is developed based on a convolutional neural network (CNN) to segment the bone surface from cross-sectional US scans. Afterwards, the segmented results are post-processed to refine the bone surface. The CNN model was trained using US images required from multiple subjects under different acquisition settings. The method has been demonstrated that it can outperform existing feature-based methods both in terms of precision and recall [77]. The method with trained model was thankfully accessible from the authors implemented in ImFusionSuite (ImFusion GmbH, Munich, Germany).

The results of applying the CNN based segmentation method on in-vivo radius bone images (B-mode) are depicted in Fig. 2.3. As it can be seen in Fig. 2.3, the US image acquired with $\theta_t = 0°$ is more contrast compared with the image obtained with a tilted ($\theta_t = 15°$) probe orientation (brighter than the surrounded soft tissues as denoted by red arrows). Thereby, the corresponding detection result for $\theta_t = 0°$ is considered better. Due to the acoustically in-homogeneous properties of bone structures, the US waves are scattered away rather than reflected back to the probe if θ_t is not zero [84]. When increasing θ_t, the ability of a US probe to properly contact the tested surface deteriorates due to the high stiffness of bone structures, which further degrades image quality in the end.

To quantitatively assess the quality of bone detection, the detected bone feature coefficient (DBFC) λ_c is defined. Intuitively, λ_c describes the true positive rate adapted to the case where the detection is confined to a curve, whereas the annotation is volumetric. It is computed by dividing the number of the true positive detected pixels over the difference in the column indeces between the leftmost and rightmost annotated pixels. A larger λ_c means to a larger true positive rate, which benefits a potential registration. The DBFC is computed using Eq. (2.1).

$$\lambda_c = \frac{|\Omega|}{G_r - G_l} \quad \text{where } \Omega = \{p | p \in \mathbf{DET} \cap \mathbf{GT}\} \tag{2.1}$$

where G_l and G_r represent the column index of the leftmost and rightmost pixels of the ground truth; **DET** and **GT** represent the sets containing all pixels belonging to the detected bone surface (line) and the labelled ground truth (area) and Ω is the intersection of **DET** and **GT**.

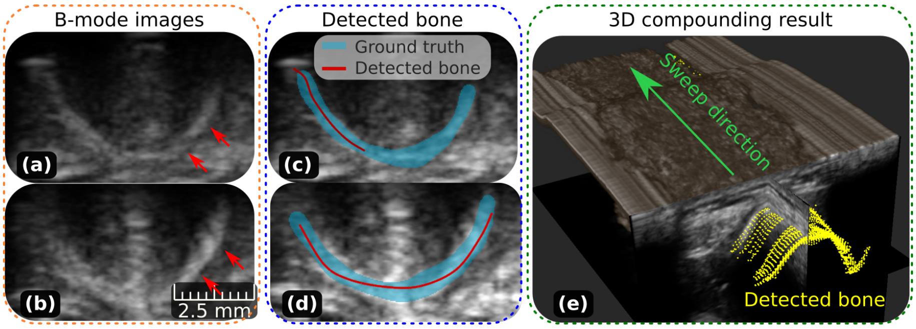


Fig. 2.3. Bone detection results for an US sweep of the radius bone for various probe orientations. (a) and (b) are the B-mode US slices from two sweeps of a volunteer's radius bone when the probe is placed in a tilted ($\theta_t = 15°$) and normal orientation ($\theta_t = 0°$). (c) and (d) show the corresponding bone detection results. (e) shows the compounded volume with the detected bone surface on each slice of the sweep for $\theta_t = 0°$. *Reprint from [9] with permission of IEEE.*

2.3.2 Detection Quality with Different Probe Orientations

To investigate the impact of θ_t on the US imaging quality and bone detection results, five US sweeps (each includes 195 2D B-mode images) were recorded from a volunteer's forearm with different θ_t ($0°$, $5°$, $10°$, $15°$, and $20°$) using a linear US probe. The forearm surface is seen as flat during short sweeps ($50\ mm$). To guarantee the constant contact force F_c, all sweeps were obtained using the compliant control method described in [85]. Furthermore, to ensure that the imaged area is the same despite the varying tilt angles, only the middle part of each sweep is considered. The bone surface detection results for the all sweeps are depicted in Fig. 2.4.

From Fig. 2.4, the sweep recorded under the normal direction ($\theta_t = 0°$) contains more B-mode scans (70%) with $\lambda_c \geq 0.75$ compared with the sweeps corresponding to the other four orientations. It can be seen that the second highest peak in that bin corresponds to the θ_t of $10°$. However, when we considering the neighbor bin ($\lambda_c \in [0.5, 0.75)$), the sweep with $\theta_c = 5°$ contains 73% of scans with $\lambda_c \geq 0.5$, whereas the sweep with $\theta_c = 10°$ contains only 66%. For sweeps with $\lambda_c < 0.25$, the sweeps with θ_t of $0°$ and $5°$ comprised only 6% and 3%, respectively. With an increase in θ_t to $20°$, the amount of sweeps with $\lambda_c < 0.25$ increased to 18%, 34%, and 29% for θ_t of $10°$, $15°$, and $20°$, respectively. Thereby, it can be seen that the bone surface extraction results improves, as the US probe is oriented closer to the normal direction.

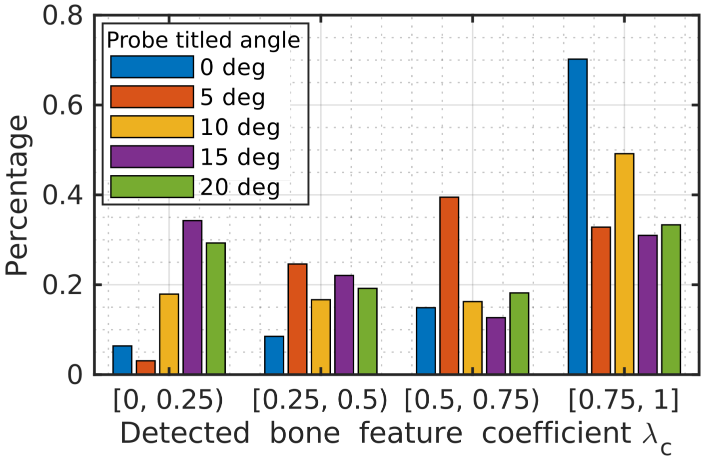


Fig. 2.4. Bone detection results for US sweeps obtained from a volunteer's forearm using a linear US probe. The probe was tilted in various angles. The contact force is $5\ N$. *Reprint from [9] with permission of IEEE.*

2.4 Normal Positioning Based on Image and Force

2.4.1 Pipeline and Theoretical Justification

The overview of the proposed approach for normal positioning of a US probe is described in this section. The normal direction of the contact surface $\mathbf{N}$ is determined based on US images and the estimated Cartesian force from joint torques. The proposed approach does not require any prior knowledge of the object. Then, an analysis of the estimated force is give to depict its dynamic behavior.

Pipeline for Computing the Normal Direction

In this work, a convex US probe (C5-2/60, Ultrasonix, Richmond, Canada) providing large penetration depth is attached to the end-effecor of a robotic manipulator (KUKA LBR iiwa 7 R800, KUKA Roboter GmbH, Augsburg, Germany).

In order to automatically align the probe axis $\mathbf{A}$ with the normal $\mathbf{N}$ from an arbitrary contact point (see Fig. 2.5 b)), the in-plane and out-of-plane alignments are carried out separately. Fig. 2.5 c) and d) describe the two adjustments from the starting orientation (black dashed line) to the desired orientation (red dashed line), respectively. Since the obtained US images are coplanar with the field of the view, the resulting image quality is highly related to the angular difference between $\mathbf{N}_i$ and $\mathbf{A}_i$. Approaches working directly with pixel intensities are sensitive to artifacts, such as shadows produced by anatomies (see red ellipse in Fig. 2.6 a)). Building upon the confidence map strategy [61], a method is originally used to compute the weighted barycenter of the regions containing objects, as opposed to those obscured by shadows (i.e. out of contact). This allows computing the correction factor for the orientation in-plane component $\mathbf{A}_i$.

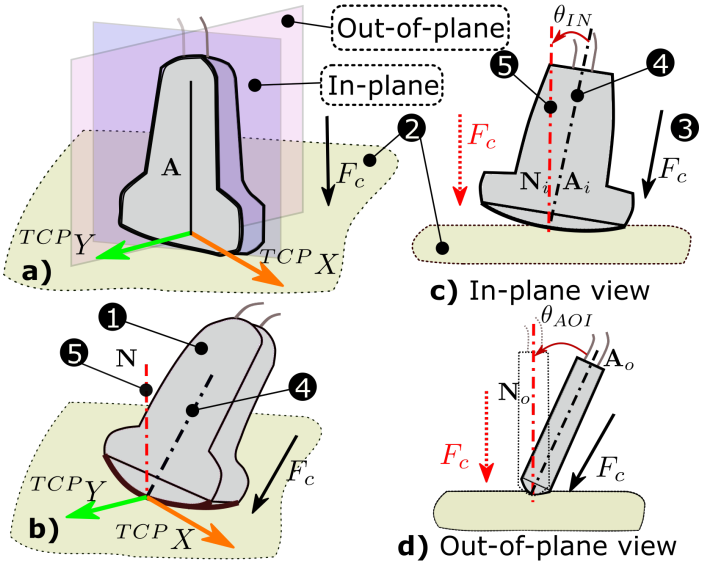


Fig. 2.5. Impact of the probe orientation on US images. 1: US probe; 2: Surface; 3: Contact force F_c; 4: probes centerline $\mathbf{A}$; 5: Normal direction $\mathbf{N}$. TCP represents the tool center point. a) and b) describe ideal and non-ideal probe orientation in 3D, respectively; c) and d) are the in-plane and out-of-plane view, respectively. *Reprint from [11] with permission of IEEE.*

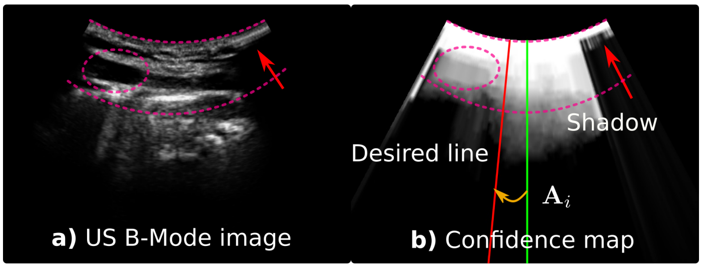


Fig. 2.6. In-plane orientation adjustment strategy. a) B-mode scan of forearm; b) corresponding confidence map. The red line is ideal $\mathbf{N}_i$. *Reprint from [11] with permission of IEEE.*

Since US images correspond by definition to the in-plane view, making the use of image-based metrics to feature the out-of-plane characteristic is challenging. To further compute the normal component $\mathbf{N}_o$, contact force is employed. Such a force is invariant to registration or calibration errors and it can be estimated by the torque sensors present in the joints. The contact force F_c between probe and patient is usually set to $3-15\ N$ in clinical practice. Reducing F_c yields a worse contact, whereas a larger F_c deteriorates the image quality because of deformation.

Then, a fan motion ($\theta_{AOI}=[-15,15°]$) around the probe long axis (orange arrow in Fig. 2.5 a)) is performed in both forward and backward directions (see Fig. 2.7 a) and b)). The estimated contact force in the direction ${}^{TCP}Y$ (green arrow in Fig. 2.5 a)) ${}^{TCP}\hat{F}_y$ is measured. The measured forces has been shown in Fig. 2.7 d) and e) for the forward and backward procedures, respectively. Besides, the theoretical ${}^{TCP}F_y$ (see Fig. 2.7 c)) can be calculated as Eq. (2.2).

$$ {}^{TCP}F_y = -F_c\, tan(\theta_{AOI}) \tag{2.2} $$

Since F_c is constant during the fan motion, ${}^{TCP}F_y$ only depends on θ_{AOI}. Ideally, θ_{AOI} is zero, which corresponds to $\mathbf{A}_o$ coinciding with $\mathbf{N}_o$. This leads to a resulting force ${}^{TCP}F_y$ of $zero$. This enables the estimation of the normal direction of unknown object based on US images and force cues.

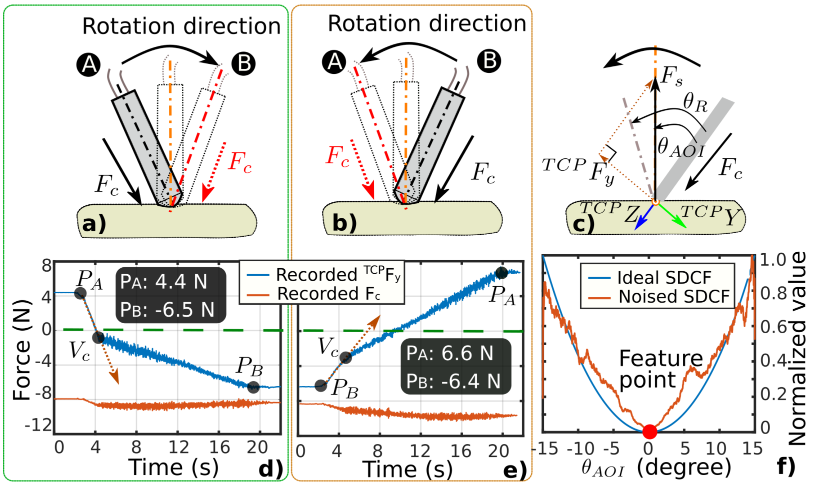


Fig. 2.7. Cartesian force recorded during fan motion. a) and b) are fan motions of the probe in the out-of-plane direction. Forward from A to B and backward from B to A; c) is the physical force model; d) and e) are the recorded ${}^{TCP}\hat{F}_y$ (blue line) and contact force (F_c, orange line) corresponding to the cases a) and b) on the gel-wax phantom. The desired force was set $F_c = 10\ N$ and the stiffness was $500\ N/m$. f) demonstrates the ideal normalized SDCF and normalized first detail coefficient of the noisy SDCF. P_A and P_B marked in d) and e) are the forces recorded at location A and B in a) and b), respectively. *Reprint from [11] with permission of IEEE.*

Cartesian Force Analysis

The external force ${}^{TCP}F_y$ at the tool center point (TCP) can be estimated from the joint torques using the dynamic model of robotic manipulator. Since the estimated Cartesian force depends on the joint configuration, singularities should be avoided in order to obtain accurate measurements. The recorded ${}^{TCP}\hat{F}_y$ during a forward and backward fan motion is shown in Fig. 2.7 d) and e). As it can be seen, the same real world point (e.g. A) can yield different recorded forces. The largest difference observed was $5\ N$ without external loads. This difference is mainly caused by the torque sensors inaccuracy and varying static friction in joints. Besides, although the forward and backward fan motions are symmetric in Cartesian frame, the joint configurations may be different because the robotic arm is redundant.

It can been seen from Fig. 2.7 d) and e), the recorded F_c is initially less than the desired F_c (10 N) because it has been partly compensated by the virtual force caused by the probe displacement in ${}^{TCP}Z$ direction. For each contact, the contact conditions are different and hence the initial recorded $\hat{F}_c$ is varying. However, after the fan motion starts, the absolute value of the recorded $\hat{F}_c$ gradually increases towards the desired force until the tissue deformation creates a proper reaction force to balance $\hat{F}_c$ applied by the robot. The corresponding time stamp is marked as t_c in Fig. 2.7. Since the initial $\hat{F}_c$ varies, t_c is also varying. Therefore the value of ${}^{TCP}F_y$ at t_c (marked as V_c) is non-deterministic. Hence, it is not possible to use a fixed threshold to determine the normal $\mathbf{N}_o$. For example, the ${}^{TCP}\hat{F}_y$ was $zero$ at about $4\ s$, whereas the normal direction is close to $12\ s$ (middle of time period) in Fig. 2.7 d).

Smooth Derivative of Contact Force (SDCF)

In order to overcome the non-deterministic behaviour of the estimated force, a new feature — smooth derivative of contact force (SDCF) — is defined as in Eq. (2.3):

$$SDCF(n) = \frac{1}{N} \sum_{i=n-N+1}^{n} |{}^{TCP}\hat{F}_y(i+1) - {}^{TCP}\hat{F}_y(i)| \tag{2.3}$$

where $n \geq N$, N is the span of the averaging filter and is empirically set to 100.

Since F_c is fixed during the fan motion, the SDCF only depends on the angle θ_{AOI}. The normalized SDCF ($n = 1$) obtained under ideal conditions is shown in Fig. 2.7 f). Further considering that F_c slightly varies during the fan motion, uniform noise between $[-1, 1]$ N was added to the desired force. Then, the noisy SDCF is computed using the noisy F_c with $n = 100$. The normalized first detail coefficient of the Discrete Wavelet Transform (DWT) (wavelet: sym4, level: 5) result of noisy SDCF is also depicted in Fig. 2.7 f).

It can be seen from the figure, both ideal and detail coefficient of noisy SDCF are minimal (marked with a red point), when the probe's axis $\mathbf{A}$ coincides with the surface normal $\mathbf{N}$. This means that the SDCF is a promising feature to identify the normal direction of unknown object surface, even when $\hat{F}_c$ slightly differs from the desired F_c. Besides, since the SDCF represents the change rate of ${}^{TCP}\hat{F}_y$ rather than the absolute force value, the negative influence caused by the non-deterministic behavior in the beginning is reduced. Furthermore, an approach based on finding the proper minimum of the SDCF is likely to yield better results than a threshold-based approach, especially when non-stationary objects are involved.

2.4.2 Optimization of In-Plane Orientation

Ultrasound Confidence Map

Since US images are coplanar with the field of view, an image-based method can be used to identify the in-plane normal $\mathbf{N}_i$. The US confidence map [61] provides an objective, stable and repeatable pixel-wise assessment of imaging quality based on the estimation of the signal loss. The US images are represented by probabilistic maps $C \in \mathbb{R}^2 \rightarrow [0, 1]$, where the maximum value (1, white) means perfect signal quality and the minimum value (0, black) means no signal. For each pixel, the confidence value of C is interpreted as the chance of the signal reaching the probe face, which makes C locally more continuous than the intensity values of B-mode images (see red ellipse in Fig 2.6). Thereby, the regions where the probe is not in contact with the patient are easily recognized due to their low confidence values (see the shadow in Fig. 2.6). Thus, the US confidence map is employed here to optimize the in-plane orientation, especially in the presence of shadows at the periphery of the US scan.

Orientation Computation for In-Plane Component

The present framework is based on the use of convex US probes. For a convex probe, US images can be defined in polar coordinates as $\Omega = [r_{min}, r_{max}] \times [\theta_{min}, \theta_{max}]$, where Ω represents the US scan; r_{min} and r_{max} denote the radius of the probe and the scan depth, respectively; θ_{min} and θ_{max} are the angular offsets from the two sides of the central line ($\theta_{max} = -\theta_{min}$), respectively. The line from the US focal point and the confidence-weighted barycenter ζ_c is a good approximation of $\mathbf{N}_i$[62]. Hence the robot is moved such that $\mathbf{A}_i$ intersects ζ_c, which can be calculated as follows:

$$\zeta_c = \frac{1}{\aleph_c}\int_{\theta_{min}}^{\theta_{max}}\int_{r_{min}}^{r'}\frac{\theta}{\theta_{max}-\theta_{min}}C(\theta, r)drd\theta \tag{2.4}$$

where $\aleph_c = \int\int_{(\theta,r)\in\Omega} C(\theta, r)drd\theta$ is the accumulated confidence value over the entire image (Ω), $r' \in (r_{min}, r_{max}]$ defines the region of interest, θ is the angular deviation from central scan-line in the image and θ_{min} and θ_{max} were set to -0.5 and 0.5 respectively.

To reach the desired position, a rotational offset R_c is applied to the current position. R_c is calculated as follows:

$$R_c = -\frac{1}{2}k_c\zeta_c\Psi_p \tag{2.5}$$

where k_c is the control gain, Ψ_p is the angular field of view of the convex probe. This is done iterative until no further improvement is reached.

2.4.3 Optimization of Out-of-Plane Orientation

This section proposes a method to compute $\mathbf{N}_o$ using force values measured during a fan motion of the robotic arm. Due to the non-ignorable noise of the recorded forces, a series of de-noising steps are applied. The overview of the de-noising procedures is Fig. 2.8.

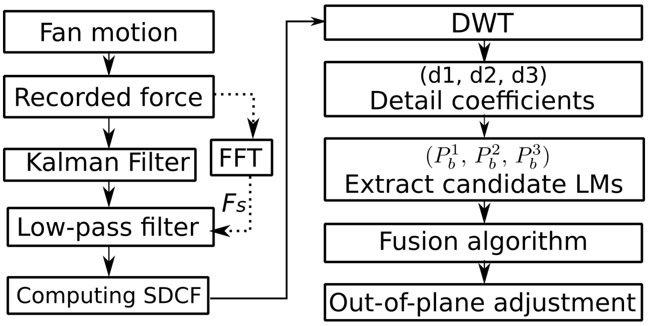


Fig. 2.8. Overview over the out-of-plane optimization algorithm. The FFT was only executed for determine the parameters of low pass filter once.

Kalman Filter Implementation

Most general-purpose robotic manipulators have the ability to estimate ${}^{TCP}F_y$ based on joint torques. However, the forces involved in RUSS are smaller than $10\ N$ due to safety concerns. Hence, digital noise has non-ignorable impact on the estimated Cartesian force for robotic

US imaging. Besides, changes of the contact point between probe and patient due to patient movement further reduce the accuracy of the recorded force (blue line in Fig. 2.9). The Kalman filter is a standard technique to reduce noise by correcting new measurements based on previous ones. Thereby, a Kalman filter is implemented to denoise the recorded force as in [86].

The fan motion is performed with a low velocity ($0.025\ rad/s$) compared to the sampling frequency of the force ($100\ Hz$). Thus the computed force values differ only slightly for consecutive sampling points. Besides, the Kalman filter operates only on the recorded force; thereby, the transition matrices for the estimated state (A) and measured state (H) are set to identity. In addition, the estimation and measurement noise covariance matrices are empirically set to 10^{-8} and 4×10^{-6}, respectively. The performance of the Kalman filter when applied to the recorded force values during two forward and backward fan-motions with the same motion parameters are described in Fig. 2.9, one on a phantom and the other one on a volunteer's lower back. The volunteer is breathing, which causes significant fluctuations of the recorded force as shown in Fig. 2.9 b).

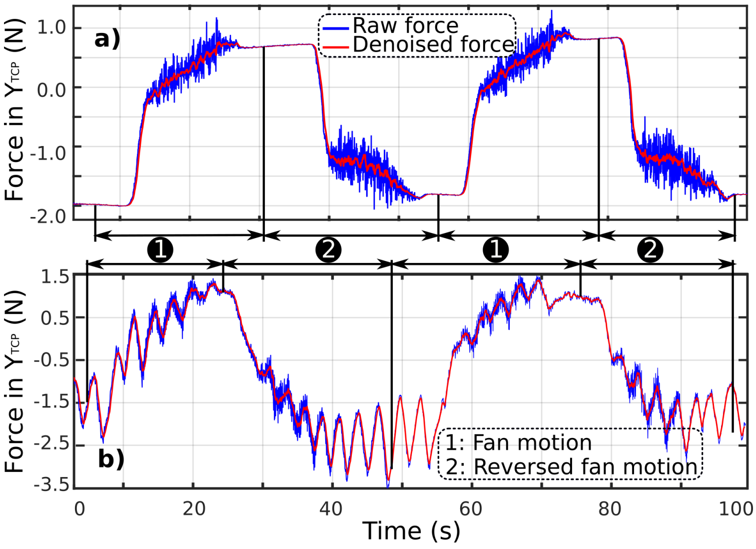


Fig. 2.9. Result of the Kalman filter. a) and b) show the experiments preformed on the phantom and volunteer's back with breathing. The involved forward and backward fan motion are performed under $F_c = 8\ N$ for the phantom and $6\ N$ for the volunteer back, stiffness$= 250\ N/m$ and Θ_{AOI} was $(-15°, 6°)$ and $(6°, -15°)$, respectively. *Reprint from [11] with permission of IEEE.*

Extraction of Frequency-Band Components from SDCF

Due to the low rotational velocity, the important force components are concentrated in the low frequency domain. Hence a low pass filter is employed after the Kalman filter to further reduce noise. In order to determine a suitable stop-band frequency Fs, the Fast Fourier Transform (FFT) is used to investigate the frequency decomposition of the recorded $^{TCP}\hat{F}_y$. An example of the FFT result and the corresponding power spectral density (PSD) of the force recorded on the phantom is shown in Fig. 2.10 a). The high frequency part can be removed because it mainly corresponds to digital noise of the sensors. Based on Fig. 2.10 a), the PSD is rapidly attenuated after $5\ Hz$ and the energy (cumulative PSD) of the signal components at $15\ Hz$ occupies 85% of the total signal energy. Thereby, Fs is set to $15\ Hz$. The result of the low-pass filter is presented in Fig. 2.10 b).

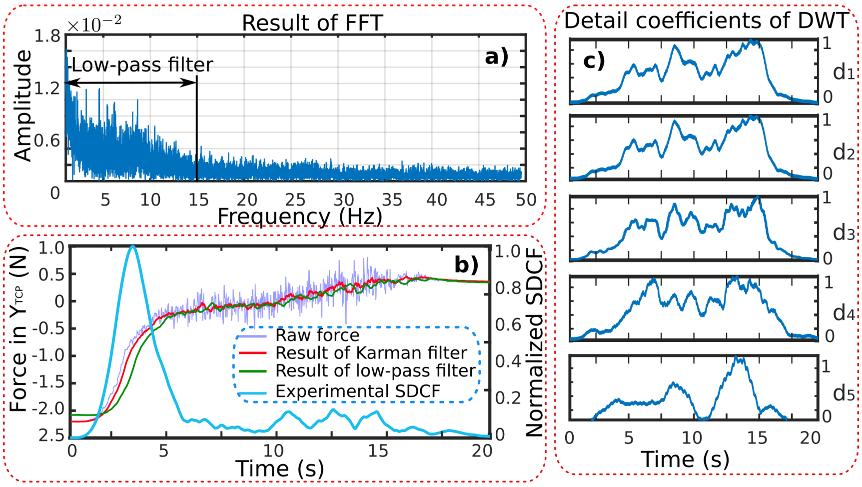


Fig. 2.10. Frequency decomposition of SDCF. a) result of FFT applied to ${}^{TCP}\hat{F}_y$; b) result of low-pass filter and the normalized SDCF. The raw force is shown with a purple line, the result of the Kalman filter and the result of the low-pass filter are depicted by red and green lines, and the blue line depicts the SDCF calculated from the result of the low-pass filter. c) Normalized detail coefficients of SDCF's DWT result. The level was set to 5, and the sym4 wavelet was used as basis. *Reprint from [11] with permission of IEEE.*

The SDCF is calculated using Eq. (2.3) based on the result of the low-pass filter. Its normalized result is shown in Fig. 2.10 b). Due to the non-periodic nature of the force caused by the fan motion, there is no identifiable main frequency in the FFT result. In order to obtain accurate estimation results, DWT is used to divide the SDCF into its different frequency band components.

The DWT provides high frequency and temporal resolution, which makes it suitable for processing unstable signals. The DWT requires a decomposition level and wavelet function as input. The sym4 wavelet function was empirically chosen as basis function in this paper. To investigate the effect of the different frequency parts, the decomposition level was set to 5. This results in five detail coefficients $d_i, i = 1 \ldots 5$ covering the intervals $[\frac{Fs}{2}\text{–}Fs]$, $[\frac{Fs}{4}\text{–}\frac{Fs}{2}]$, $[\frac{Fs}{8}\text{–}\frac{Fs}{4}]$, $[\frac{Fs}{16}\text{–}\frac{Fs}{8}]$ and $[\frac{Fs}{32}\text{–}\frac{Fs}{16}]$, respectively. The d_i have different frequency interval size, hence their amplitude range also differs. In order to balance the impact of different frequency bands on identification of the normal direction, the details signals are normalized to $[0,\ 1]$ (see Fig. 2.10 c)). Since the normalization is linear, it does not affect the locations of local minima.

It can be seen from Fig. 2.10 c), the amplitude graphs for d_4 and d_5 are unlike the other three. This is due to them having a low frequency value and small frequency range. Low frequencies are influenced by contact point changes, e.g. due to breathing. Hence, only d_1, d_2 and d_3 are used to detect $\mathbf{N}_o$.

Extracting Local Minima from SDCF

In theory, the SDCF is minimal when $\mathbf{A}$ is aligned with $\mathbf{N}$ as shown in Fig. 2.7 f). Due to the noise, this minimum is not necessarily the global minimum for real scenarios. The experimental detail signals $(d_i, i = 1, 2, 3)$ from Fig. 2.10 c) have multiple local minima (LM). The LM can be detected by comparing nearby points using Eq. (2.6). The detected LM are depicted as red stars in Fig. 2.11.

$$I_i = \{t |\ d_i(t-\Delta t) > d_i(t) \wedge d_i(t+\Delta t) > d_i(t)\} \tag{2.6}$$

where I is the set used to save the corresponding time stamps of LM, $i = 1, 2, 3$ is the detail coefficient index, t is the time stamp and $\Delta t = 10$ ms is the sampling interval. The equation is written in set-builder notation.

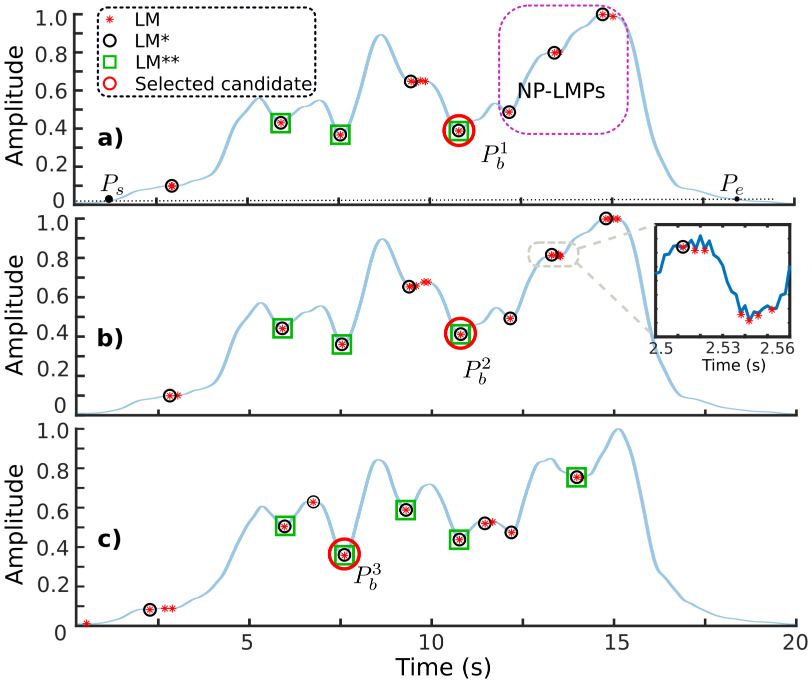


Fig. 2.11. Suitable local minimum selection. a), b) and c) correspond to d_1, d_2 and d_3, respectively. The parameters are set as follows: $W_1 = 0.8$, $L_1 = 0.04$, $H_1 = 0.08$. *Reprint from [11] with permission of IEEE.*

Then, a method is proposed to detect the right LM for each d_i. During this process, two variables (LM* and LM**) are defined. LM* are the LM remaining after removing neighboring LM (LM located in close proximity to each other, see the enlarged plot in Fig. 2.11 b). LM** are the subset of LM* after further removing the so-called NP-LMs (noise-perturbed) which are caused by noise perturbing the estimated force (see the pink rectangle in Fig. 2.11).

To extract the LM*, a sliding window filter with width W_1 and length L_1 is applied to remove all but the first element for each neighboring LM group. The LM* are marked with a black circle in Fig. 2.11 and the corresponding timestamps form the set I_i^*. To further remove NP-LM, the input signal d_i is partitioned into segments along the LM*. If a LM* amplitude is close to the maximum amplitude of one of its neighbouring segments, it is discarded. The remaining LM (LM**) are marked with a green square in Fig. 2.11. A set I_i^{**} containing their corresponding timestamps is created. This is explained in Eq. (2.7).

$$I_i^{**} = \{t |\ {}^{\mathrm{L}}M_i^k - d_i(t) > H_1 \wedge {}^{\mathrm{R}}M_i^k - d_i(t) > H_1\} \tag{2.7}$$

where $i = 1, 2, 3$ is the detail coefficient index, t are the recorded time stamps, $k = 1,\ 2, \ldots,\ length(I_i^*) - 2$, H_1 is the amplitude threshold, ${}^{\mathrm{L}}M_i^k$ and ${}^{\mathrm{R}}M_i^k$ are the maximum amplitude values of the left and right segments adjacent to the kth LM. The equation is written in set-builder notation.

In order to identify the desired LM candidate for each d_i, a feature A^* is defined as in Eq. (2.8).

$$A^*(LM, i) = ({}^{\mathrm{L}}M_i^{LM} - d_i(t)) \cdot ({}^{\mathrm{R}}M_i^{LM} - d_i(t)) \tag{2.8}$$

where $i = 1, 2, 3$ is the detail coefficient index, LM is the LM, $t = I_i^{**}(k)$ is the timestamp of the LM, ${}^{\mathrm{L}}M_i^j$ and ${}^{\mathrm{R}}M_i^j$ are the maximum amplitude values of the segment left and right of LM.

For each d_i the LM** with the largest A^* is selected as the final candidate (depicted with red circle in Fig. 2.11).

Computation of Normal Direction

The potential normal direction is computed from the final LM (P_b^1, P_b^2 and P_b^3) of the detail coefficients d_i, respectively. To further improve the robustness of the identified result, a fusion algorithm using the three candidate LM is proposed. As it can be seen from Fig. 2.12, there are three possible cases and five potential results:

Case 1: The maximum A^* of P_b^1, P_b^2 and P_b^3 is larger than the threshold H_A. The timestamp of the LM with maximum A^* is returned.

Case 2: The three selected LM have a similar timestamp. Then the mean timestamp is returned.

Case 3: Two of the selected LM have similar timestamps. The mean of their timestamps is denoted as t_{mean}, while the third timestamp t_d differs. The two selected LM with similar timestamps are denoted as a and b, the last one is called c. T_1 and T_2 are computed as described in Case 3.1 and Case 3.2, respectively. If T_2 is greater than T_1, t_d is returned, otherwise t_{mean}.

Case 3.1: A LM corresponding to t_{mean} is searched in the set LM$_c^{**}$. If there is an element in I_c^{**} with a timestamp not more than H_3 seconds away from t_{mean}, the search is considered a success. T_1 is computed as the sum of the A^* of the searched LM (if available), P_b^a and P_b^b.

Case 3.2: As in case 3.1, the two corresponding LM for t_d are searched among the LM$_a^{**}$ and LM$_b^{**}$. T_2 is computed as the sum of the A^* of the searched LM (if available) and P_b^c.

Failure: If all three selected LM have different timestamps, then the algorithm returns a failure.

Finally, the returned $\mathbf{N}_o$ is the $\mathbf{A}_o$ corresponding to the returned timestamp.

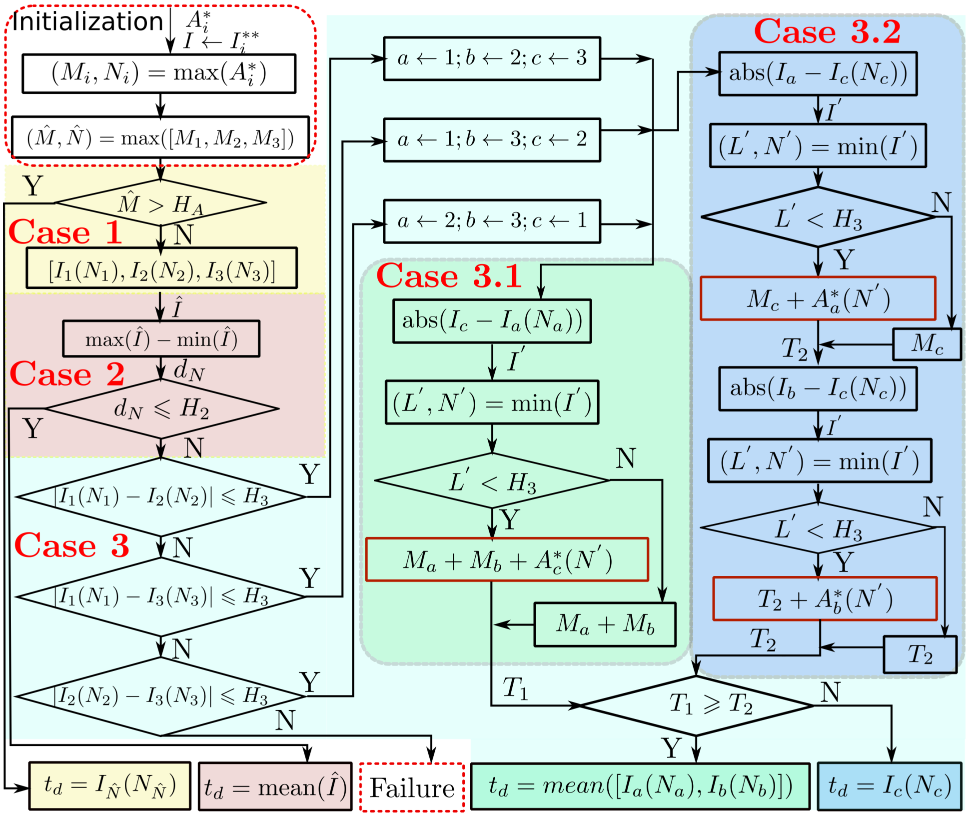


Fig. 2.12. Fusion algorithm to extract $\mathbf{N}_o$ from the recorded probe orientations based on timestamps. The parameters are set as follows: $H_A = 0.3$, $H_2 = 0.8$ and $H_3 = 1.5$. *Reprint from [11] with permission of IEEE.*

2.4.4 Results

Experimental Setup

The overall experimental setup has been depicted in Fig. 2.13. A convex US probe is attached to the robot flange using a 3D-printed fixture. The B-mode US images are acquired using an Ultrasonix RP machine (BK Ultrasound, Peabody, MA, USA). The images are sent to a computer using an Ethernet cable and the OpenIGTLink protocol. The computer communicates with the robot using Robot Operating System (ROS). Control commands and the robot status are exchanged at 100 Hz. The algorithm was evaluated on a gel-wax phantom with a flat surface and on a human volunteer (lower arm, upper arm, lower back). The soft model is fixed in the foam plastic, imitating the function of skin, to avoid greatly deformation during the scanning. The phantom is placed on a flat surface. Hence the ground truth is represented as ${}^{B}\vec{V}_g = [0, 0, 1]$ in the robotic base frame. For the human volunteer, the ground truth is computed based on scanning the surface around the desired path. The angular difference between estimated normal $\hat{\mathbf{N}}$ and ground truth normal $\mathbf{N}$ is computed using the equation: $\theta = cos^{-1}(\hat{\mathbf{N}} \cdot \mathbf{N} / (|\hat{\mathbf{N}}| \cdot |\mathbf{N}|))$.

Validation of Estimating $\mathbf{N}_i$ using the Gel-Wax Phantom

The computation of $\mathbf{N}_i$ is the first step of the proposed algorithm. It was validated using a series of experiments carried out on the gel-wax phantom using three different initial θ_{AOI}

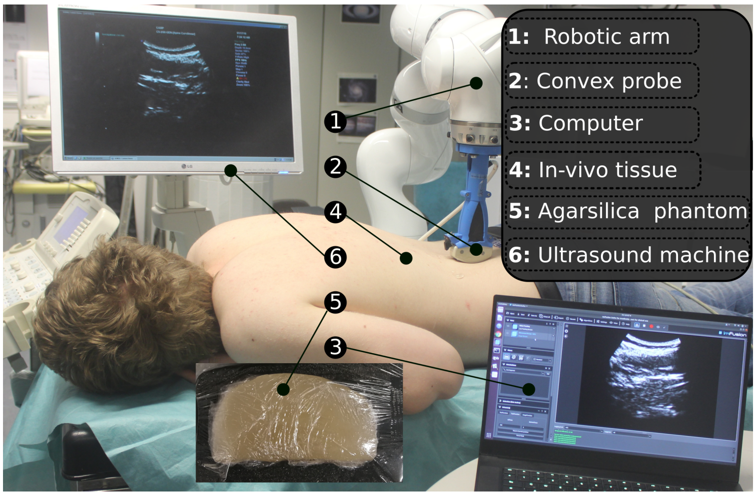


Fig. 2.13. Experimental setup. *Reprint from [11] with permission of IEEE.*

and four different initial offset angles θ_{IN} in the in-plane view, where θ_{IN} is the angular difference between $\mathbf{A}_i$ and $\mathbf{N}_i$. The US confidence map is calculated from US images, which are highly sensible to deformations. In order to assess the influence of the deformation on in-plane adjustment, the robot is controlled in position mode. Each of the twelve experiments is performed four times with different depths ($z = z_0 + 0.5, z_0 + 1, z_0 + 1.5, z_0 + 2\ mm$), where z_0 is the depth at which the probe starts touching the phantom.

$\mathbf{N}_i$ is estimated using Eq. (2.4) with $r' = r_{min} + 0.2(r_{max} - r_{min})$. The error ($e_{in}$, mean±STD) for different θ_{IN} and θ_{AOI} is shown in TABLE 2.1. The experiments show that the method is able to achieve good estimates for $\mathbf{N}_i$.

Tab. 2.1. Mean Absolute Error (±STD) of In-Plane Identified Results

e_{in} \ θ_{AOI} / θ_{IN}	0	5	10
0	1.11 ± 0.55	0.18 ± 0.09	1.20 ± 0.79
3	0.94 ± 0.66	0.84 ± 0.29	0.56 ± 0.36
6	1.38 ± 0.41	1.61 ± 0.53	1.55 ± 0.39
9	2.71 ± 0.57	2.92 ± 0.20	2.69 ± 0.81

*Units are in degree.

Validation of Estimating $\mathbf{N}_o$ using the Gel-Wax Phantom

Five experiments were performed on the gel-wax phantom with flat surface and $F_c = 8\ N$. To make the tests more realistic and challenging, an asymmetric fan motion ($\theta_{AOI} = [-6, 15°]$) is carried out. The calculated angular differences θ_c between $\mathbf{A}$ and $\mathbf{N}$ of the five paired fan motions are shown in Fig. 2.14. It is calculated by $\sqrt{\theta_{IN}^2 + \theta_{AOI}^2}$.

$\mathbf{A}_i$ is not necessarily aligned with $\mathbf{N}_i$, therefore even if $\mathbf{A}_o$ is aligned with $\mathbf{N}_o$, an error of zero is not necessarily achieved. Therefore the minimum error for each fan motion is equal to θ_{IN}. For multiple repeated fan motions, it can be seen that θ_{IN} increases with the number of fan motions. This is caused by the probe sliding on the contact surface, due to the US gel and deformations of the soft tissue. The $\mathbf{A_o}$ corresponding to the timestamps of the computed results (red and green dots in Fig. 2.14) are taken as estimated $\mathbf{N}_o$. Except for one result (the third fan motion, green dot), all results are close to the desired direction ($< 5°$), despite θ_{IN} being non-zero. Furthermore, five of the ten results are very close to the minimum ($< 0.5°$). This demonstrates that the method can detect $\mathbf{N}_o$ purely based on the estimated force provided by a robot without external force sensor.

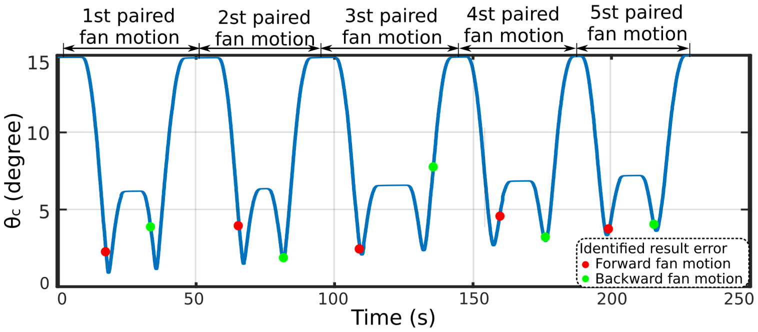


Fig. 2.14. Computed results for five paired fan motions. The red and green dots depict the results for the forward and backward fan motion, respectively. *Reprint from [11] with permission of IEEE.*

To further analyse the proposed method, the normalized amplitude for the d_i for the fifth paired fan motion (largest initial angle difference with $3.3°$ and $3.6°$) are shown in Fig. 2.15.

The final differences between the ideal $\mathbf{N}$ and detected $\hat{\mathbf{N}}$ are $3.5°$ and $3.8°$ for the forward and backward fan motion. However, when the in-plane component is ignored, the errors are $1.0°$ ($\sqrt{3.47^2 - 3.32^2}$) and $1.3°$ ($\sqrt{3.84^2 - 3.60^2}$), respectively. This means that $\mathbf{N}_o$ can be estimated reliably, even if the estimation for $\mathbf{N}_i$ is poor.

Comparison with Human Operators

To compare the proposed algorithm with US operators, six volunteers experienced in US scanning were asked to manually position the US probe along the normal direction of the same phantom ten times (60 times total). The average errors for the human operators are $3.2 \pm 1.7°$ while the error of the proposed method on the same phantom is $3.1 \pm 1.0°$. It can been seen from Fig. 2.16, the operators perform better than the proposed method in the best case (error: $0.2°$ vs $1.7°$) while the error of proposed method is more stable and in the worst case ($3.6°$) performs better than human operators ($5.9°$). Based on a t-test (probability $p = 0.52 > 0.05$), there is no significant difference between the proposed method and human operators. Hence the proposed method can compute $\mathbf{N}$ with a comparable accuracy to the participating human operators on a phantom.

In-Vivo Validation

The recorded force is highly related to the properties of the scanned tissue. In order to evaluate the applicability of the proposed method on humans, experiments were performed on the forearm, upper arm and lower back with normal breathing of a volunteer. Eight scans were

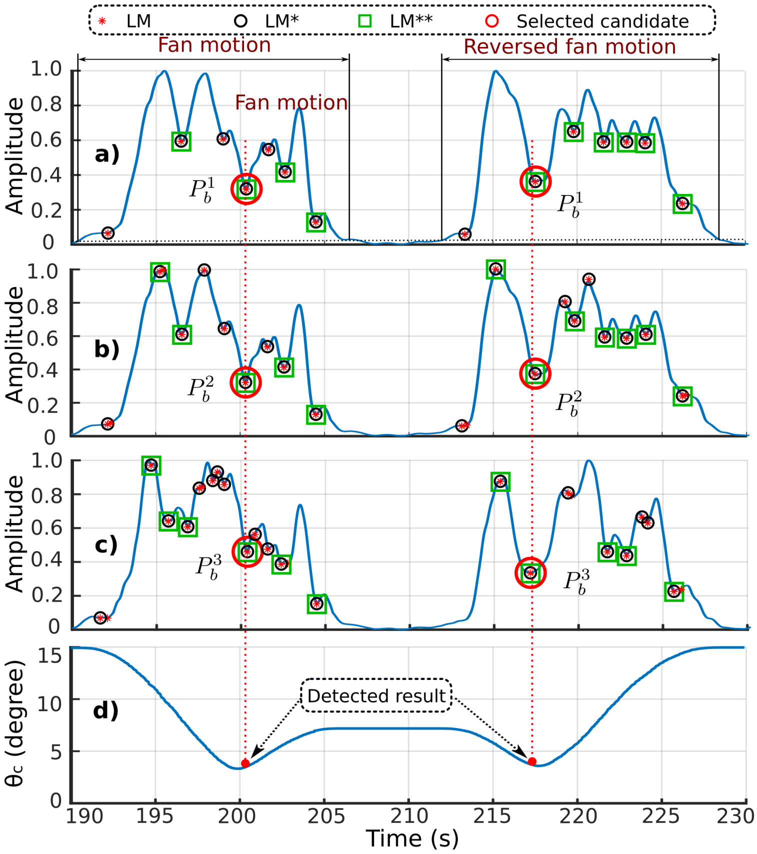


Fig. 2.15. Feature curve for one paired fan motion. The two red dashed lines represent the selected timestamps based on the detail signals of SDCF (d_i) for the forward and backward fan motion, respectively. *Reprint from [11] with permission of IEEE.*

performed per anatomy. The computed errors is depicted in Fig. 2.16. Since there is variation in the involved tissue stiffness, the robot stiffness was set to $250\ N/m$, $350\ N/m$ and $450\ N/m$ during the tests to investigate the impact of the robot stiffness and $F_c = 6\ N$.

It can be seen from Fig. 2.16 that the proposed algorithm performed best on the phantom. This is due to the phantom being stationary whereas humans exhibit motion. For the forearm, the lowest error for each set of experiments is smaller than $2°$, and the 25th percentile of all three sets of experiments is close to $2°$. The algorithm performed best when the stiffness was $450\ N/m$, having a maximum error of $5.8°$. For the upper arm, the average errors corresponding to the various stiffness are distributed around $5.5°$. For stiffness $350\ N/m$, both the variance and the upper error bound ($7.6°$) were the lowest among the varying stiffness values. The error for the lower back is larger than for the other anatomies due to respiratory motion. However, the algorithm still achieves good best case results ($1.7°$, $2.9°$ and $1.7°$) for the different stiffness, respectively. The results suggest that the robot stiffness should be adapted to the tissues stiffness, in order to accurately estimate the normal direction of local surface.

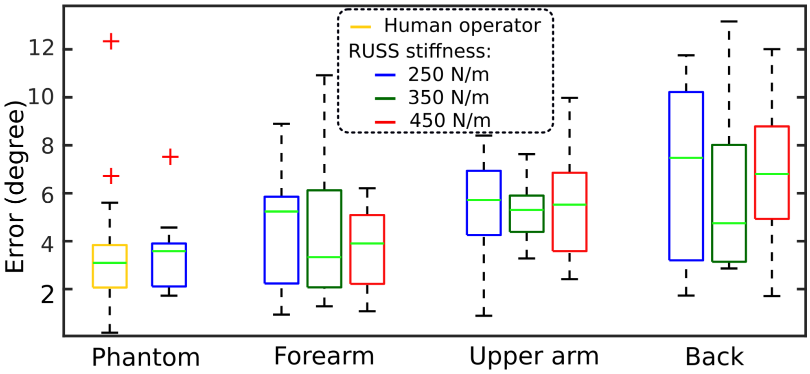


Fig. 2.16. The errors of the estimated normal direction with phantom/volunteer. The first two plots are the experiments performed on the phantom by humans and the proposed method. The other plots show the results of the method on in-vivo tissues. *Reprint from [11] with permission of IEEE.*

2.5 Normal Positioning Based on Mechanical Model

2.5.1 Normal Direction Detection

Strategy for Searching the Normal Direction

To ensure the imaging quality and patient safety, the compliant control scheme described in Section 2.5.2 is employed to maintain a constant force F_c along the probe centerline **C** during US scans (see Fig. 2.17). The target tissues in orthopedic applications (i.e. limbs) are rigid due to the presence of bones, whose elastic modulus is more than $8.91\ Gpa$ [87]. Thus, when a constant F_c is exerted along the probe, reaction force $\mathbf{F}_r$ is generated along **N** of the constraint surface to balance F_c. The other reaction force component normal to F_c varies according to the change in the relative posture between probe **C** and unknown **N** as shown in Fig. 2.17 (d). According to the concept of reaction force defined in theoretical mechanics, it can be inferred that when an external force F_c is applied along the normal direction of a contact surface **N**, the resulting reaction force components in the other two orthogonal directions should be zero.

To estimate **N** of an unknown contact surface, two rotations with a given point of contact around two orthogonal axes (x_{sa} and y_{sa}) are executed. During the process of rotation, the reaction force component ${}^{sa}F_y$ and ${}^{sa'}F_x$ (see Fig. 2.17 (c) and (d)) are measured by a F/T sensor attached onto the flange, respectively. Then, the components of **N** lying in the plane defined by the rotation axis ($\mathbf{N}_x$ or $\mathbf{N}_y$) is estimated. After each rotation, the robot is moved to align the probe center line **C** with the estimated direction.

Desired $\mathbf{N}_x$ is located inside the virtual plane $\mathbb{A}$ that comprises desired **N** and the randomly selected axis $\mathbf{x}_{sa}$. Desired $\mathbf{N}_x$ is approximated by **C** when the probe is rotated on the plane $\mathbb{A}$. Theoretically, this is also the posture resulting in ${}^{sa}F_y$ to be zero. Then, the second rotation around the axis $\mathbf{y}_{sa}$ orthogonal to $\mathbf{x}_{sa}$ is conducted. It is used to estimate $\mathbf{N}_y$ where ${}^{sa'}F_x$

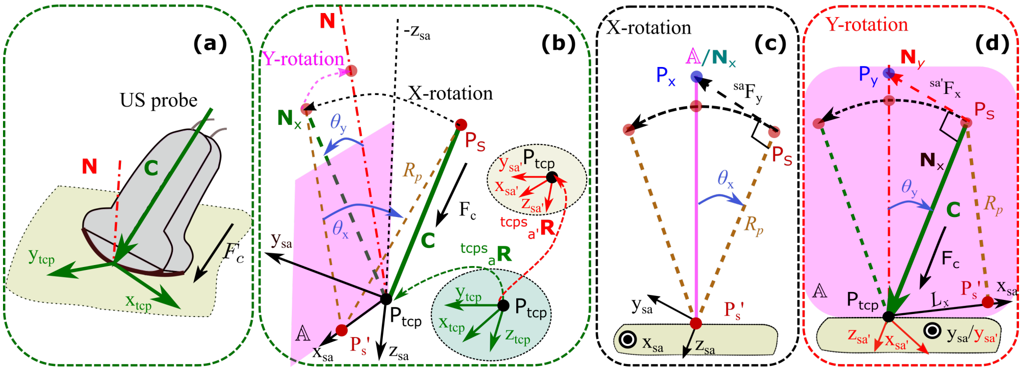


Fig. 2.17. Theoretical mechanical model of the two successive rotations; (a) a 3D view of a random contact between the convex probe and the unknown constraint surface; (b) the 3D model of two orthogonal searching rotations around random $\mathbf{x}_{sa}$ and corresponding $\mathbf{y}_{sa}$ at a fixed point, respectively; (c) 2D view of the rotation around randomly selected axis $\mathbf{x}_{sa}$; (d) 2D view of the rotation around the computed axis $\mathbf{y}_{sa}$, that is normal to $\mathbf{x}_{sa}$. Desired $\mathbf{N}_x$ is estimated based on the X-rotation. Then the Y-rotation is performed from the estimated $\mathbf{N}_x$. Computed $\mathbf{N}_y$ is used to approximate $\mathbf{N}$ of the unknown constraint surface. P_s and P_{tcp} denote the F/T sensor frame $\{s\}$ and TCP frame $\{tcp\}$, respectively. *Reprint from [9] with permission of IEEE.*

is reduced to zero. The two rotations around $\mathbf{x}_{sa}$ and $\mathbf{y}_{sa}$ are denoted as X-rotation and Y-rotation, as shown in Fig. 2.17 (c) and (d).

However, human tissue, such as the human limbs, are not completely rigid. There are some soft tissues (i.e. muscles) between the skin and an underlying bone. In real scenarios, small and unexpected deformations of soft tissues are generated to balance the external force. Thus, the recorded force is affected, particularly, in the applications requiring the applied force to be small. Accordingly, to identify the normal direction of human tissues, the measured forces ${}^{sa}F_y$ and ${}^{sa'}F_x$ (see Fig. 2.17 (c) and (d)) will not be exactly zero when the US probe is accurately placed along desired direction $\mathbf{N}_x$ or $\mathbf{N}_y$. To address the limitation of the simple threshold-based method, here a model-based optimization method is employed to extract $\mathbf{N}$. Since the model-based method estimates the desired direction based on the whole experimental force data rather than on a single value, it is more robust in terms of adapting to the small differences between the ideal and real cases.

Mechanical Model

In this subsection, we explain the theoretical mechanical model describing two successive rotations (X-rotation and Y-rotation) with a given contact point. The goal of the mechanical model is to establish a framework that allows understanding the exact relationship between the measurements of the F/T sensor and the tilt angle of a probe. The problem can be decomposed into two independent subproblems. First, $\mathbf{N}_x$ is estimated using the rotation around $\mathbf{x}_{sa}$. Then, $\mathbf{N}_y$ is computed based on the rotation around $\mathbf{y}_{sa}$, after the X-rotation alignment has been executed.

First, the F/T sensor origin $\mathrm{P_s}$ is projected onto the selected rotation axis $\mathbf{x}_{sa}$ yielding point $\mathrm{P_s^{'}}$. The equivalent radius of X-rotation R_p is equal to the distance between $\mathrm{P_s}$ and $\mathrm{P_s^{'}}$.

$$R_p = \sqrt{|\mathbf{C}|^2 - L_x^2} \tag{2.9}$$

where $\mathbf{C}$ is the probe length vector (including the length of fixture) from $\mathrm{P_s}$ to the TCP P_{tcp}, and L_x is the distance between $\mathrm{P}_\mathrm{s}^{'}$ and P_{tcp}.

After placing the US probe on a given contact point with a random posture, L_x is computed using Eq. (2.10):

$$L_x = \mathbf{C} \cdot \mathbf{x}_{sa} \tag{2.10}$$

where $\mathbf{x}_{sa}$ is a unit vector.

Then, a unit vector $\mathbf{z}_{sa}$ that is orthogonal to $\mathbf{x}_{sa}$ is defined in the direction of the equivalent radius R_p as shown in Fig. 2.17 (c). The unit vector $\mathbf{z}_{sa}$ is calculated using Eq. (2.11):

$$\mathbf{z}_{sa} = \frac{\mathbf{C} + L_x \cdot \mathbf{x}_{sa}}{|\mathbf{C} + L_x \cdot \mathbf{x}_{sa}|} \tag{2.11}$$

Accordingly, the Cartesian coordinate frame {sa} corresponding to the randomly selected $\mathbf{x}_{sa}$ is well-defined by $\mathbf{x}_{sa}$, $\mathbf{z}_{sa}$, and $\mathbf{y}_{sa} = \mathbf{z}_{sa} \times \mathbf{x}_{sa}$. The reaction force component in $\mathbf{y}_{sa}$ (${}^{sa}F_y$) is able to be represented by a function with respect to the rotation angle around $\mathbf{x}_{sa}$ (θ_x) from plane $\mathbb{A}$ to real-time equivalent radius R_p.

$${}^{sa}F_y = \mathrm{sgn}(\theta_x)\frac{R_p}{|\mathbf{C}|}\, F_c \cdot \tan\theta_x \tag{2.12}$$

where $\mathrm{sgn}(\theta_{out})$ is the sign function.

The measured ${}^{sa}\hat{F}_y$ is computed based on the recorded force using Eq. (2.13):

$$\begin{aligned} {}^{sa}\hat{\mathbf{F}}_r &= {}^{sa}_{tcp}\mathbf{R}\, {}^{tcp}_{s}\mathbf{R}\, {}^{s}\mathbf{F}_r \\ {}^{sa}\hat{F}_y &= {}^{sa}\, \mathbf{F}_r(y) \end{aligned} \tag{2.13}$$

where ${}^{A}_{B}\mathbf{R}$ is the rotation matrix used to describe the frame {B} relatively to frame {A}; ${}^{sa}_{tcp}\mathbf{R} = [{}^{tcp}\mathbf{x}_{sa},\ {}^{tcp}\mathbf{y}_{sa},\ {}^{tcp}\mathbf{z}_{sa}]^{-1}$, ${}^{tcp}_{s}\mathbf{R}$ is fixed after attaching the F/T sensor on the flange.

Upon the successful completion of the X-rotation computation, the probe is placed along the computed $\hat{\mathbf{N}}_x$, which is located at the plane $\mathbb{A}$ in the ideal case. Then, the corresponding Y-rotation around $\mathbf{y}_{sa}$ is executed. To intuitively describe the reaction force with respect to the rotation angle around $\mathbf{y}_{sa}$ (θ_y) ranging from desired $\mathbf{N}_x$ to real-time probe $\mathbb{C}$, the other Cartesian coordinate frame {sa'} can be defined, where $\mathbf{y}_{sa'} = \mathbf{y}_{sa}$, $\mathbf{z}_{sa'} = \mathbf{C}$ and $\mathbf{x}_{sa'} = \mathbf{y}_{sa'} \times \mathbf{z}_{sa'}$ (Fig. 2.17 (d)). The reaction force component in $\mathbf{x}_{sa'}$ (${}^{sa'}F_x$) is computed using Eq. (2.14):

$${}^{sa'}F_x = -\mathrm{sgn}(\theta_y)\, F_c \cdot \tan\theta_y \tag{2.14}$$

In addition, the measured ${}^{sa'}\hat{F}_x$ during the Y-rotation is calculated based on the sensor output using Eq. (2.15):

$$\begin{aligned} {}^{sa'}\hat{\mathbf{F}}_r &= {}^{sa'}_{tcp}\mathbf{R}\, {}^{tcp}_{s}\mathbf{R}\, {}^{s}\mathbf{F}_r \\ {}^{sa'}\hat{F}_x &= {}^{sa'}\, \mathbf{F}_r(x) \end{aligned} \tag{2.15}$$

where ${}^{sa'}_{tcp}\mathbf{R} = [{}^{tcp}\mathbf{x}_{sa'},\ {}^{tcp}\mathbf{y}_{sa'},\ {}^{tcp}\mathbf{z}_{sa'}]^{-1}$.

The reaction force model of searching rotations around a random axis ($\mathbf{x}_{sa}$) is formulated in Eq. (2.12), and Eq. (2.14). Since the real-time Cartesian frame {tcp} can be computed based on the robotic kinematic model, it is deemed a suitable candidate to be used as a rotation axis without the need for additional computations. Moreover, since the US probe structure (both linear and convex) is symmetrical around the long axis ($\mathbf{x}_{tcp}$) and short axis ($\mathbf{y}_{tcp}$) (see Fig. 2.20), $\mathbf{x}_{tcp}$ and $\mathbf{y}_{tcp}$ are the center lines of the contact area in two orthogonal directions when the probe is close to the normal direction of the constraint surface. This means that the center contact lines remain unchanged during the process of searching rotations, respectively. The unchanged centerline of the contact area indicates that no additional deformation is generated along the rotation axis. Therefore, the two searching rotations can be executed around $\mathbf{x}_{tcp}$ and $\mathbf{y}_{tcp}$. Then, the theoretical reaction force model defined in Eq. (2.12) and Eq. (2.14) can be rewritten as Eq. (2.16), and Eq. (2.17) for the orthogonal out-of-plane rotation ($\mathbf{x}_{tcp}$) and in-plane rotation (co-planar with the US image plane, $\mathbf{y}_{tcp}$), respectively.

$$^{tcp}F_y = \mathrm{sgn}(\theta_{out})\ F_c \cdot \tan\theta_{out} \tag{2.16}$$

where θ_{out} is the out-of-plane offset angle between **N** and **C**. Here, θ_{out} is identical to θ_x when $\mathbf{x}_{sa} = \mathbf{x}_{tcp}$.

$$^{tcp}F_x = -\mathrm{sgn}(\theta_{in})\ F_c \cdot \tan\theta_{in} \tag{2.17}$$

where θ_{in} is the in-plane offset angle between **N** and **C**. Here, θ_{in} is identical to θ_y when $\mathbf{y}_{sa} = \mathbf{y}_{tcp}$.

As a result of comparing Eq. (2.16) and Eq. (2.17), it can be seen that both forces corresponding to the out-of-plane and in-plane searching rotations are able to be computed using the general equation described in Eq. (2.18):

$$^{tcp}F_i = F_a \cdot \tan\theta_j + F_b \tag{2.18}$$

where $(i, j) = (y, out)$ or (x, in); F_a and F_b are constant variables for each rotation. Here, F_b is introduced to reduce the negative effect caused by the force bias related to the initial contact condition for the real experimental data.

Extracting the Normal Direction

Based on the recorded reaction force $^{tcp}\hat{\mathbf{F}}_r$ and the previously derived mechanical model $^{tcp}F_i$, $i = x$ or y (see Eq. (2.18)), a model-based orientation optimization method is proposed to obtain first $\mathbf{N}_o$, and then $\mathbf{N}_i$ as a result of the out-of-plane and in-plane rotation.

However, since the derivative of a tangent monotonically increases until infinity, a small variation in $^{tcp}\hat{\mathbf{F}}_r$ may lead to non-negligible errors in a tangent-based fitting result. However, the required maximum angular error in this task is less than $20°$ ($|\theta_i| < 20°$). Therefore, the tangent-based equation (Eq. (2.18)) derived from the theoretical mechanical model can be approximated by a sine-based function applying small-angle approximation ($20° = 0.35$ rad). Compared with the tangent-based equation, the derivative of the sine-based function is limited to the range $[-1, 1]$. Thus, it is more robust to fit the sine-based function to the noise-perturbed $^{tcp}\hat{\mathbf{F}}_r$ using Eq. (2.19):

$$ {}^{tcp}F_i = p_1^i + p_2^i \ \sin\left(2\pi p_3^i t + p_4^i\right) \tag{2.19} $$

where $\theta_j = 2\pi p_3^i t + p_4^i$, $(i, j) = (x, out)$ or (y, in), and t is the time stamp of the recorded force data.

The unknown parameters in Eq. (2.19) are estimated based on the experimental fluctuation-compensated reaction force (FCRF) ${}^{tcp}\hat{\mathbf{F}}_r$ using the constrained non-linear optimization problem described in Eq. (2.20).

$$ \begin{aligned} & min \ \frac{1}{2N} \sum_{j=1}^{N} [{}^{tcp}F_i(j) \ - \ {}^{tcp}\hat{F}_i(j)]^2 \\ & s.t. \ \begin{cases} 2\pi p_3^i t(k) + p_4^i \leq \eta_{max} \\ 2\pi p_3^i t(k) + p_4^i \geq \eta_{min} \end{cases} \quad \forall \ k \in [1, N] \end{aligned} \tag{2.20} $$

where $i = x$ or y; N is the length of ${}^{tcp}\hat{F}_r$; η_{max} and η_{min} are the boundaries limiting the search space, respectively.

Eq. (2.20) can be solved using the Sequential Quadratic Programming (SQP) method. The desired direction for each of rotations is at $\theta_j = 0$. Therefore, the normal ($\mathbf{N}_o$ or $\mathbf{N}_i$) is approximated by the recorded orientation **R** corresponding to the time stamp t_f making $\theta_j = 0$ using Eq. (2.21).

$$ \mathbf{N}_d = \mathbf{R}(t_f) \quad \text{when } \theta_j = 0 \tag{2.21} $$

The fitted result and the corresponding feature curve used to identify $\mathbf{N}_o$ for an example of the out-of-plane rotation are shown in Fig. 2.18. Here, a representative experiment determining the out-of-plane normal is performed. Then based on the optimized time stamp t_f computed by Eq. (2.21), the ideal $\mathbf{N}_o$ is approximated by the orientation encoded by the recorded probe orientation $\mathbf{R}(t_f)$. The angular error between the ground truth $\mathbf{N}_o$ and $\mathbf{R}(t_f)$ is $0.7°$ (error$=\arccos \frac{\mathbf{N}_o \ \cdot \ \mathbf{R}(t_f)}{|\mathbf{N}_o| \ |\mathbf{R}(t_f)|}$).

2.5.2 Control Architecture and Data Processing

In this section, we first describe the control architecture. Then, several de-noising procedures are introduced to mitigate the presence of the large noise in the raw force data recorded during the out-of-plane or in-plane rotations. An overview of the algorithm is shown in Fig. 2.19.

Compliant Control Architecture

The experimental setup for a gel-wax phantom with a planar surface is illustrated in Fig. 2.20. To measure the interaction force during the estimation of the normal direction **N**, an F/T sensor was attached to the robotic flange. To validate the proposed method in different cases, a phantom with a flat surface is placed onto the tilted surfaces with varied tilt angles. In

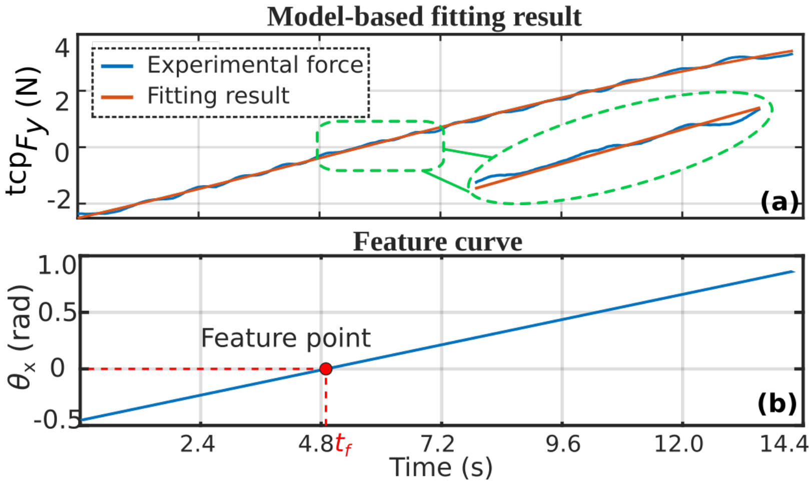


Fig. 2.18. (a) Fitting result. (b) feature point identification. The experimental force is the denoised force data recorded by F/T sensor. Parameters for the out-of-plane part: rotation range $\theta_{out} \in [10°, -15°]$, phantom tilt angle $\alpha_t = 5°$ and initial in-plane position $\theta_{in} = -10°$. *Reprint from [9] with permission of IEEE.*

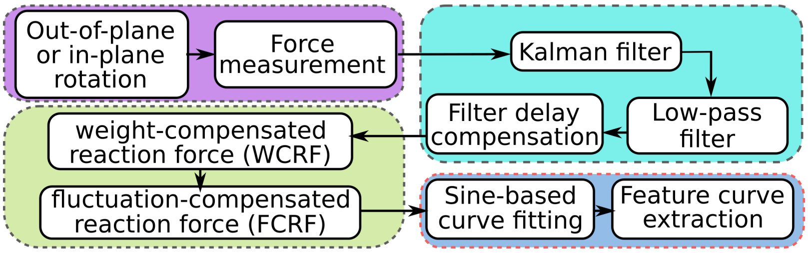


Fig. 2.19. Workflow of estimating the normal direction based on force measurement. *Reprint from [9] with permission of IEEE.*

addition, since the rotations are performed around $\mathbf{x}_{tcp}$ and $\mathbf{y}_{tcp}$ (frame {tcp}) fixed on the probe, the symbols $\mathbf{N}_x$ and $\mathbf{N}_y$ used in the rotation around an arbitrary frame {sa} are replaced by $\mathbf{N}_o$ and $\mathbf{N}_i$, respectively.

The robot motion is controlled via a 1-DOF compliant controller combined with a 5-DOF position controller as in [85]. The acting 5-DOF position controller is used to execute the desired scan trajectory or to limit the movement of the contact point around the given position. The compliant controller is applied to exert the constant desired contact force ($\mathbf{F}_d$) between the US probe and tissue aiming to guarantee the US imaging quality and patient safety during screening. The position controller is used to execute the desired scan trajectory. The spring-like compliant controller is represented in Eq. (2.22).

$$^{tcp}\mathbf{F} = \mathbf{F}_d - \mathbf{K}(\mathbf{P}_i - \mathbf{P}_0) \tag{2.22}$$

where $^{tcp}\mathbf{F} \in \mathbb{R}^m$ is the real force acting onto the contact point in the frame {tcp} (tool center point); $\mathbf{K} \in \mathbb{R}^{m\times 1}$ is the stiffness vector; $\mathbf{P}_0$ is the position where the robot enters force mode; $\mathbf{P}_i$ is the current position.

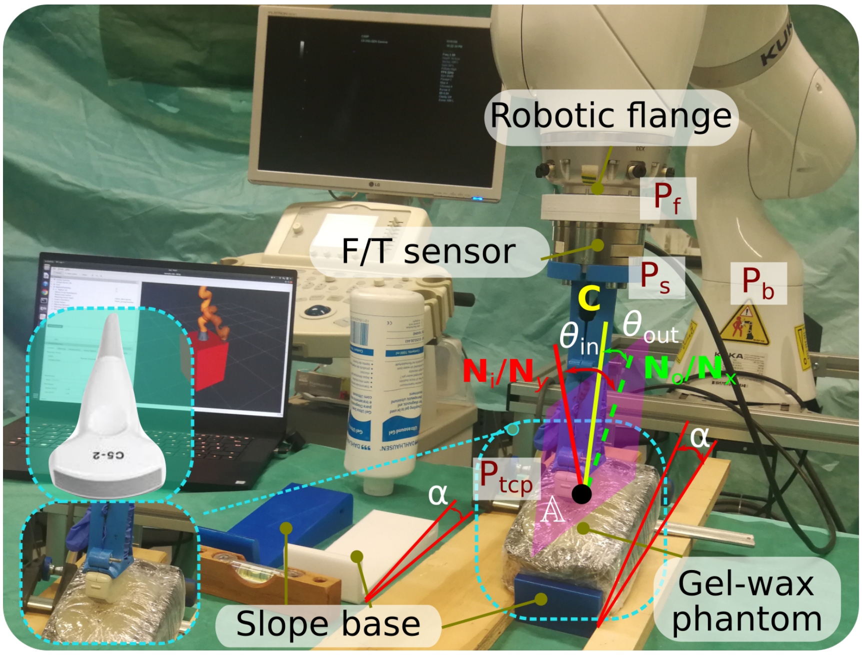


Fig. 2.20. Experimental setup on the gel-wax phantom. The rotation is performed around $\mathbf{x}_{tcp}$ (out-of-plane view) and $\mathbf{y}_{tcp}$ (in-plane view) rather than the frame {sa} corresponding to an arbitrary axis. The symbols $\mathbf{N}_x$, $\mathbf{N}_y$, θ_x and θ_y in Fig. 2.17 are replaced with $\mathbf{N}_o$, $\mathbf{N}_i$, θ_{out} and θ_{in}, respectively. P_b, P_f, P_s and P_{tcp} denote the origin coordinates in the robotic base frame $\{b\}$, the robotic flange frame $\{f\}$, the F/T sensor frame $\{s\}$, and TCP frame $\{tcp\}$, respectively. Different slope bases are used to tilt the phantom with different angles ($\alpha_t = 0^\circ$, 5°, 10° and 15°). *Reprint from [9] with permission of IEEE.*

Force Calibration using Particle Swarm Optimization

To obtain the correct reaction force component (${}^{tcp}F_i$ in Eq. (2.18)) during the searching rotation, the weight of the US probe, the custom-designed holder and the preloading force (to mount the sensor on the robot) were identified and then compensated. The ouput from the F/T sensor (${}^s\mathbf{F}_m$) constituted the combined result of the real contact force (${}^s\mathbf{F}_r$), the probe weight (${}^b\mathbf{W}$), and the initial force offsets (${}^s\mathbf{F}_{off}$), including of the sensor bias and the preloading force.

$$ {}^s\mathbf{F}_m = {}^s\mathbf{F}_r + {}^s_b\mathbf{R}\, {}^b\mathbf{W} + {}^s\mathbf{F}_{off} \tag{2.23} $$

where ${}^s_b\mathbf{R} = {}^{tcp}_{s}\mathbf{R}^{-1}\ {}^{b}_{tcp}\mathbf{R}^{-1}$; ${}^{tcp}_{s}\mathbf{R} = [0\ 1\ 0;\ -1\ 0\ 0;\ 0\ 0\ 1]$ is the rotation matrix from $\{s\}$ to $\{tcp\}$; ${}^{b}_{tcp}\mathbf{R}$ is the varied transformation mapping data from $\{tcp\}$ to $\{b\}$, which is obtained from the forward kinematics; $\{b\}$ is the robot base frame; $\{s\}$ is the F/T sensor frame.

The PSO algorithm is employed to identify the unknown parameters ${}^b\mathbf{W}$ and ${}^s\mathbf{F}_{off}$ based on the experimentally recorded force data. PSO is an iterative population-based stochastic optimization technique inspired by the foraging behavior of bird flocks [88]. Its main advantages are acceptable performance in terms of finding global optima and its simple implementation [14].

The PSO utilizes a set of particles to find an optimal solution by moving the particles around in a search space. The procedure is guided by minimizing the value of the fitness function F_n. In this work, F_n is defined by Eq. (2.24):

$$F_n = \frac{1}{3N} \sum_{i=1}^{N} \| {}^{s}\mathbf{F}_m(i) - ({}^{s}_{b}\mathbf{R}(i){}^{b}\mathbf{W}(i) - {}^{s}\mathbf{F}_{off}) \| \tag{2.24}$$

where N is the number of measurements.

De-noising and Compensation of Recorded Force

The measured force data contains noise due to the undesired contact conditions and digital noise of the F/T sensor. Therefore, a set of denoising steps are applied to the measured force data. First, a Kalman Filter (KF) is employed to remove the Gaussian digital noise introduced by the F/T sensor and the robotic servoing system. Following the steps listed in [11], the transition matrices for the estimated state (A) and the measured value (H) of KF are set to be the identity matrix because the last measurement could be set as the predicted value due to the high sampling frequency ($F_s = 83Hz$) of the F/T sensor and the slow rotational velocity (0.025 rad/s). The process and measurement noise model are empirically chosen to be $\mathcal{N}(0, 10^{-4})$ and $\mathcal{N}(0, 10^{-6})$, respectively. The results of KF is depicted in Fig. 2.21.

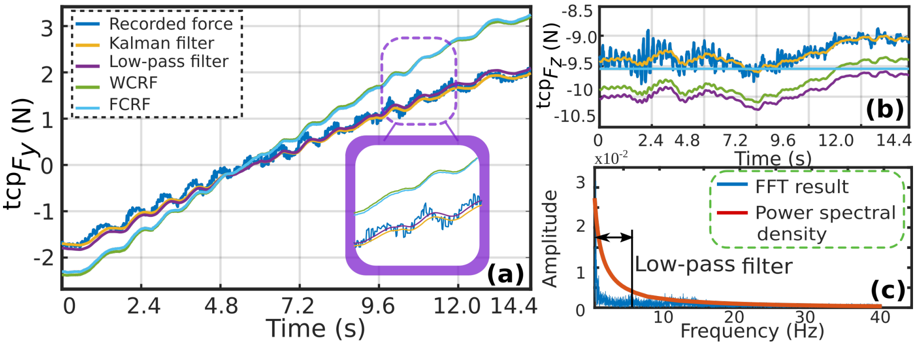


Fig. 2.21. Effect of the processing steps on the data recorded during the out-of-plane rotation; (a) and (b) display ${}^{tcp}F_y$ and ${}^{tcp}F_z$, respectively; (c) shows the FFT result on ${}^{tcp}F_y$. Experimental parameters: rotation range $[10, -15]^{\circ}$, the phantom tilt angle $\alpha_t = 5^{\circ}$, initial orientation $\theta_{in} = -10^{\circ}$ and $\theta_{out} = -15^{\circ}$. *Reprint from [9] with permission of IEEE.*

Since the rotations are non-periodic and executed with a relatively slow velocity ($0.025\ rad/s$), the resulting force recorded during the rotation is mainly distributed in the low-frequency part. Therefore, a low-pass filter is applied to restrain the high-frequency noise. Its stop-band frequency is determined using the FFT and the processed results of the measured force F_m and the corresponding power spectral density (PSD) are shown in Fig. 2.21 (c). The PSD is rapidly attenuated after $5\ Hz$, and the energy (cumulative PSD) at $5\ Hz$ occupies over 86% of the total signal energy. Hence, the stop-band frequency is set to $5\ Hz$ in this work.

Nevertheless, the force signal will experience a constant phase shift (group delay) after applying the low-pass filter. This corresponds to time delay (t_{de}), defined as the derivative of the phase with respect to frequency. More details can be found in [89]. The output of the low-pass filter are further corrected by shifting the denoised signal by t_{de}. Regarding the general S-tap low-pass filter, t_{de} is calculated using Eq. (2.25):

$$t_{de} = (S-1)/(2 * F_s) \tag{2.25}$$

To further extract the force corresponding to the searching rotation from low-pass filter result, the weight of the probe (including fixture) should be compensated. In all experiments, the bias of the force sensor measurement is compensated when the probe is aligned along the ${}^{b}z$ direction. Then, based on the probe weight (${}^{s}\mathbf{F}_w$ and ${}^{s}\mathbf{F}_{off}$) estimated by PSO, the weight-compensated reaction force (WCRF) in frame $\{tcp\}$ (${}^{tcp}\mathbf{F}_r$) is calculated by rewriting Eq. (2.23) as Eq. (2.26).

$${}^{tcp}\mathbf{F}_r = {}^{tcp}_{s}\mathbf{R}({}^{s}\mathbf{F}_m - {}^{s}_{b}\mathbf{R}\,{}^{b}\mathbf{W} - {}^{s}\mathbf{F}_{off}) \tag{2.26}$$

It can be seen from Fig. 2.21 (b) that the actual contact force along **C** (${}^{tcp}F_z$) is not exactly equal to the desired F_c, due to inhomogeneous deformations and small slides happened between the probe and the contact surface. To limit the influence of small fluctuations of recorded ${}^{tcp}F_z \in (F_c - \Delta F, F_c + \Delta F)$, the average value of ${}^{tcp}F_z$ recorded during the rotation is computed as ${}^{tcp}\overline{F}_z$. Thus, the fluctuation-compensated reaction force (FCRF) ${}^{tcp}\hat{\mathbf{F}}_r = [{}^{tcp}\hat{F}_x, {}^{tcp}\hat{F}_y, {}^{tcp}\hat{F}_z]$ is calculated by mapping the fluctuated ${}^{tcp}F_z$ to ${}^{tcp}\overline{F}_z$ using Eq. (2.27).

$${}^{tcp}\hat{\mathbf{F}}_r(i) = \frac{{}^{tcp}\overline{F}_z}{{}^{tcp}F_z(i)}\,{}^{tcp}\mathbf{F}(i), \quad \text{if } \frac{{}^{tcp}\overline{F}_z}{{}^{tcp}F_z(i)} < T_1 \tag{2.27}$$

where ${}^{tcp}\overline{F}_z = \frac{1}{N}\sum_{i=1}^{N}\,{}^{tcp}F_z(i)$, N is the number of measurements.

2.5.3 Results

Experimental Setup

The proposed method was evaluated using a KUKA LBR iiwa 7 R800 robot (KUKA Roboter GmbH, Augsburg, Germany), an Ultrasonic Sonix RP machine (BK Ultrasound, Peabody, USA) and a Gamma F/T sensor (ATI Industrial Automation, USA) attached to the end-effector via a custom-designed 3D printed holder. The B-mode images were acquired using both a C5-2/60 GPS convex probe and a L14-5/38 GPS linear probe (BK Ultrasound, Peabody, USA).

The robot is controlled via a software module[1] based on the ROS framework. The B-mode scans are transferred to a workstation (Intel Core i7-8750H CPU, 16 GB RAM, GeForce GTX 1060) using the OpenIGTLink protocol. The robot's status and force measurements are exchanged at $100\ Hz$ and $83\ Hz$, respectively. The gel-wax phantom with a flat surface is placed on a flat table. Hence the ground truth is represented by $[0,\ 0,\ 1]$ when $\alpha_t = 0°$. As for the phantom with a non-zero α_t ($5, 10\ and\ 15°$) and in-vivo volunteer tissues, the ground truth is represented by the normal direction of a local surface composed by three neighbor points

---

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

Tab. 2.2. Performance of Model-based Identification method on Tilted Phantom (Mean±SD)

e_{out} \ θ_{out} / θ_{in}	Forward Out-of-plane rotation						Backward Out-of-plane rotation						$\bar{e}_{off}$
	[-15, 10]	[-10, 10]	[-5, 10]	[-15, 5]	[-10, 5]	[-5, 5]	[10, -15]	[10, -10]	[10, -5]	[5, -15]	[5, -10]	[5, -5]	
-10	1.0	**0.7**	2.9	2.6	1.5	2.6	2.9	1.7	2.2	2.4	**0.3**	**0**	1.7±1.0
-5	**0**	**7.3**	**0.4**	3.8	3.4	1.9	**0.7**	2.6	2.8	3.7	2.8	2.1	2.6±1.9
0	3.5	**0.4**	2.0	**0.9**	**0.1**	2.2	**0.9**	**0.1**	2.0	2.9	**0.3**	2.0	1.2±1.0
5	3.0	**0.6**	1.4	2.1	3.4	1.7	4.0	4.4	**0.2**	**5.2**	**6.4**	**0.9**	2.8±2.0
10	**0.1**	1.8	**0.6**	1.2	2.0	2.6	2.6	2.8	2.5	2.7	2.5	2.2	2.2±0.7
$\bar{e}_{ra}$	1.5±1.6	2.1±2.9	1.4±1.0	2.1±1.1	2.1±1.4	2.2±0.4	2.2±1.4	2.3±1.6	1.9±1.0	3.4±1.1	2.5±2.5	1.4±1.0	2.1±1.5

*Tilted angle α_t = 5 deg * Units are in deg * **Green**: e_{out}<1.0 deg * **Red**: e_{out}>5.0 deg

distributed around the given position as in [11]. The neighbouring points were manually selected. We have experimentally evaluated that this method measures the normal with an accuracy of less than $0.1°$ for a gel wax phantom.

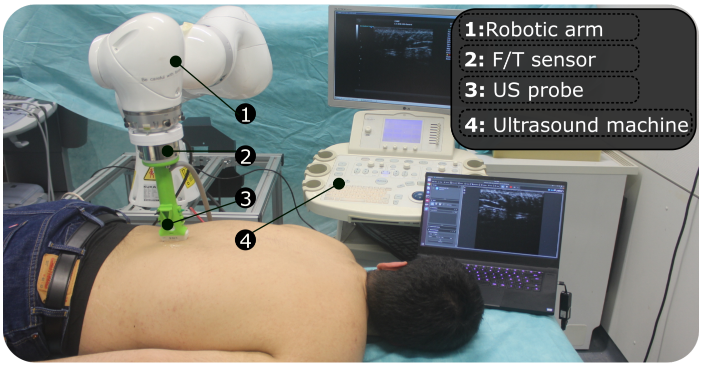


Fig. 2.22. Experimental setup. *Reprint from [9] with permission of IEEE.*

External Load Calibration Result

The probe weight ${}^b\mathbf{W}$ and the initial force offset ${}^s\mathbf{F}_{off}$ are optimized by minimizing the fitness function F_n (Eq. (2.24)) using PSO. The parameters of PSO are listed as follows: acceleration constants $c_1 = c_2 = 2$; inertia weight $\xi_{min} = 0.4$ and $\xi_{max} = 0.9$; particle size $J_p = 24$; the maximum number of iterations: 2000. The measured force (${}^s\mathbf{F}_m$) is recorded while the end-effector is rotated around x_{tcp} and y_{tcp} without any external load. Then the weight of the convex probe (${}^b\mathbf{W} = [0\ 0\ -4.33]$N) and initial force offset (${}^s\mathbf{F}_{off} = [-0.01\ 0.04\ -4.31]$N) are obtained using PSO. The algorithm terminated after 1400 iterations, returning the final F_n of 0.0022. Repeating the procedures, the probe weight and initial force offset for the linear probe are calculated as follows: ${}^b\mathbf{W} = [0\ 0\ -3.05]$N) and ${}^s\mathbf{F}_{off} = [-0.03\ -0.05\ -3.06]$N.

Robustness of the Model-Based Identification Algorithm

To evaluate the robustness of the proposed method aiming to ensure its usability in real applications without any requirements for the prior knowledge of the constraint surface, a set of out-of-plane rotations with varying initial in-plane angle offsets θ_{in} and different rotation parameters (direction, start angle, and end angle) are executed. All experiments are conducted on the same gel-wax phantom with a flat surface. The phantom is placed on a tilted surface ($\alpha_t = 5°$) (see Fig. 2.20) to mimic the non-planar human tissues. The resulting angular difference between the estimated normal $\hat{\mathbf{N}}_o$ and actual $\mathbf{N}_o$ are listed in TABLE 2.2.

It can be seen from TABLE 2.2 that the average angular difference $\bar{e}_{out}$ (mean±SD) summed over all trials is $2.1 \pm 1.5°$. This is smaller than the error of human operators reported in [11], which is $3.2 \pm 1.7°$. Furthermore, in two cases, e_{out} of zero was achieved, and in 51 out of 60 cases (85%), e_{out} was less than three degrees. The two cases with the zero error contain one sample with the largest rotation range ($[-15, 10]$) and one sample with the smallest rotation range ($[-5, 5]$). The average errors for each row and column are denoted by $\bar{e}_{ra}$ and $\bar{e}_{off}$ in the table. It can be seen that there is no significant difference between the experiments conducted using different θ_{in} and θ_{out} values. When grouped by a rotation direction, the average errors $\bar{e}_{out}$ for the forward and backward scanning are $1.9 \pm 1.5°$ and $2.3 \pm 1.5°$, respectively. However, the t-test (wiht the probability $p = 0.27 > 0.05$) yielded that the search direction has no significant impact on the identification result. Therefore, the proposed method demonstrated the capability of providing the sufficiently accurate estimate of $\mathbf{N}_o$ for various θ_{in} and rotation parameters (direction, starting angle, and ending angle).

However, there are three suboptimal cases ($e_{out} \geq 5°$, marked in red in TABLE 2.2). Two of them occurred when the start or the end of the rotation is close to the desired direction ($\leq 5°$). This is because the US probe is not a mass point as modeled in Fig. 2.20. This indicates that an asymmetric searching rotation introduces a bias of the measured force data, which is not in line with the theoretical model. Nevertheless, the largest $\bar{e}_{ra}$ is only $3.4 \pm 1.1°$ when the absolute end angle is $10°$ greater than the absolute start angle ($[5, -15]$). In addition, while starting or stopping the movement of the robotic arm, the motion is not stable during the beginning and ending phases, which will further corrupt the force measurement. This suggests that both start and end points of a rotation had to be sufficiently far from $\mathbf{N}_d$ ($\geq 5°$) to guarantee accurate identification results.

In addition, the impact of the rotation velocity on identification performance is also investigated. Besides 0.025 rad/s, three other velocities (0.015, 0.05 and 0.075 rad/s) are tested in the experiments using the same frequency for force data acquisition. For each velocity, at least four experiments are repeated under the same condition. The angular error between the detected direction and ground truth slowly grows (0.38 ± 0.23, 0.41 ± 0.28, 0.94 ± 0.40, and $1.40 \pm 0.11°$) as velocity increases from 0.015 to 0.075 rad/s. Based on the result, the error behaves close to linear with respect to the velocity. This is because a larger velocity results in larger motion during the same time period. Thus, it is possible to maintain the identification accuracy by increasing the recording frequency of the robot pose and force data when a larger velocity is used. But a slow motion means that the contact condition is closer to the static case while a fast rotation will introduce an unmodeled dynamic contact situation between the probe and human tissue. This means that too high velocity may result in a failure to find a good estimation. Therefore, the rotation velocity should be carefully selected to make a trade-off between accuracy and efficiency. If the accuracy is the main concern for a specific application, the probe is suggested to be rotated slowly.

Validation on Gel-Wax Phantom

In this section, we systemically describe the performance of the proposed method concerning both convex and linear probes on a gel-wax phantom. To fully validate the proposed method, the phantom is placed on various titled surfaces ($\alpha_t = 0, 5, 10, 15°$). The absolute angular difference for the out-of-plane and in-plane rotations are shown in Fig. 2.23.

Regarding the out-of-plane experiments, six rotations were executed for each pair of θ_{in} ($-10, -5, 0, 5$ and $10°$) and the probe type (convex and linear). Then, the rotation ranging $[-15, 10°]$ offset from the desired location is performed. Concerning in-plane evaluation, the rotation ranges from $[-13, 8°]$. For each pair of a probe type (linear, convex) and θ_{out} ($-5, 0, 5$), the rotation is repeated five times. The range of θ_{out} in this case is smaller, as the out of plane alignment is conducted before the in-plane rotation.

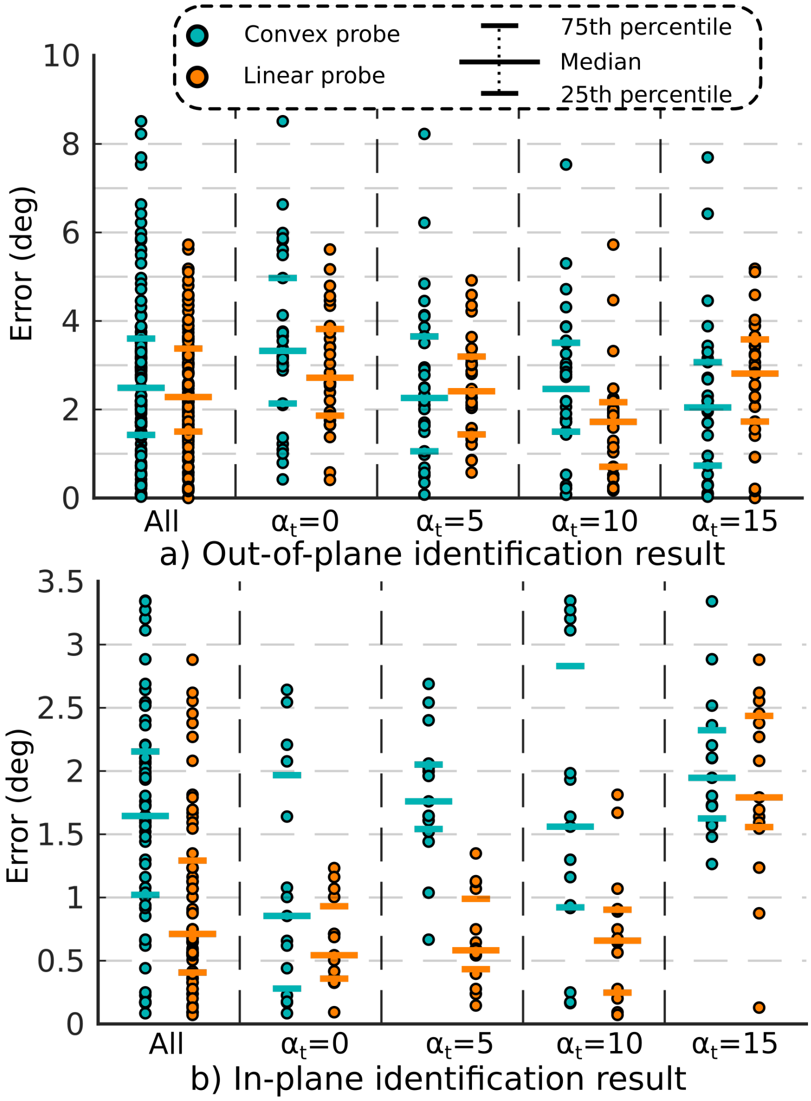


Fig. 2.23. Absolute angular difference between ground truth $\mathbf{N}_o$ and $\mathbf{N}_i$ and the estimated values $\hat{\mathbf{N}}_o$ and $\hat{\mathbf{N}}_i$ for out-of-plane and in-plane rotation, respectively. The circles represent the absolute angular difference for each experiment, the bar represents the mean value. The bottom and top short bars represent the 25th and 75th percentile, respectively. There are 30 experiments for out-of-plane and 15 for in-plane for each probe type and each α_t. Several initial angular offsets were used ($[-10, -5, 0 - 5, 10]$). Here, α_t denotes the tilt angle of the phantom. *Reprint from [9] with permission of IEEE.*

It can be seen from Fig. 2.23 (a) that the average angular difference (± SD) for the 120 out-of-plane rotations (e_{out}) is $2.7 \pm 1.9°$ for the convex and $2.5 \pm 1.3°$ for the linear probe. More than 75% of e_{out} for all 240 rotations (both convex and linear probes) are below $3.6°$. As can be seen in Fig. 2.23 (b), the average in-plane error (e_{in}) is $1.7 \pm 0.9°$ and $1.0 \pm 0.7°$ for the convex and linear probes, respectively, based on 60 rotations. The model-based optimization

Tab. 2.3. Performance of Model-based Identification method on Tilted Phantom (Mean±SD)

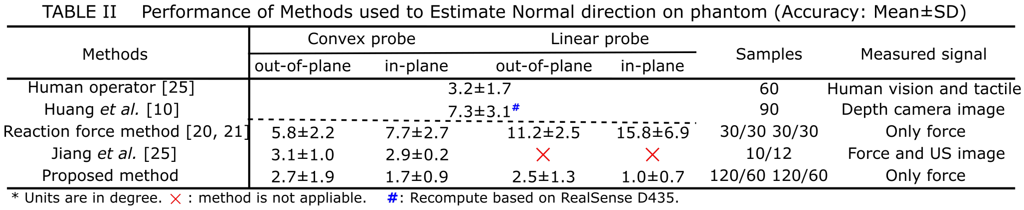


TABLE II Performance of Methods used to Estimate Normal direction on phantom (Accuracy: Mean±SD)

Methods	Convex probe		Linear probe		Samples	Measured signal
	out-of-plane	in-plane	out-of-plane	in-plane		
Human operator [25]	3.2±1.7				60	Human vision and tactile
Huang *et al.* [10]	7.3±3.1#				90	Depth camera image
Reaction force method [20, 21]	5.8±2.2	7.7±2.7	11.2±2.5	15.8±6.9	30/30 30/30	Only force
Jiang *et al.* [25]	3.1±1.0	2.9±0.2	✕	✕	10/12	Force and US image
Proposed method	2.7±1.9	1.7±0.9	2.5±1.3	1.0±0.7	120/60 120/60	Only force

* Units are in degree. ✕ : method is not appliable. #: Recompute based on RealSense D435.

also demonstrates robustness to the various tilt angles α_t used. Due to its flat tip, the proposed method achieves better performance for the linear probe rather than for the convex probe, which contains an arched tip. When considering all 360 rotations (the convex and linear probes, in- and out-of-plane), over 75% of the samples exhibites an error less than $4°$. In totally, over 75% detected results among all 360 experiments (both probe types and searching rotations) are less than $4°$.

It is also noteworthy that although the average error of the in-plane and out-of-plane optimizations are rather close to each other (less than $1°$), the maximum error of the in-plane optimization is much less compared with that of the out-of-plane optimization(convex: $8.5°$ vs 3.3°; linear: $5.7°$ vs 2.9°). The differences between the performances of the in-plane and out-of-plane optimizations are caused by the probe structure for which the length in $\mathbf{x}_{tcp}$ is much larger than the width in $\mathbf{y}_{tcp}$ direction. For example, the length and the width of the linear probe are 50 mm and 13 mm, respectively. The larger length results in greater force changes when the probe is rotated around the same angle. Thus, the in-plane optimization results are better than those of the out-of-plane optimization as shown in Fig. 2.23. In addition, the performance differences between the probes (see Fig. 2.23 (b)) are caused by a similar reason. This is because the linear tip of the linear probe makes it more sensitive to the in-plane rotation compared with the round tip of the convex probe.

In addition, to demonstrate the superiority of the proposed method, we compared it with existing approaches and also human operators as shown in TABLE 2.3. A red cross indicates that a method cannot be applied for the corresponding scenario (e.g. probe type). The performance of the camera based method [57] is computed based on the angular difference between the ground truth and the estimated normal direction of a flat calibration board at different positions and timestamps (90 data in total). In addition, a simple method based on the reaction force ($\mathbf{N} = [\frac{F_x}{|F|}, \frac{F_y}{|F|}, \frac{F_z}{|F|}]$) used in [81, 82] also does not work well in this case, particularly for the linear probe. This is mainly caused by the probe's geometry and the properties of human tissues, which are not as stiff as a workpiece. Since the probe width is much smaller than its length (linear probe: $13\ mm$ vs $50\ mm$), the out-of-plane performance is better than the in-plane performance for both linear and convex probes (see TABLE 2.3).

Then we further compared our proposed method with the state of art probe orientation optimization method [11], which is the first work aimed to fully optimize a US probe orientation (both in-plane and out-of-plane). However, the in-plane optimization method is only applicable to a convex probe and required sufficient amounts of shadows in US images. The force based out-of-plane adjustment was developed based on the searching for an optimized minimum force deviation. In the present study, all available experimental data has been taken into consideration rather than focusing on local minima [11], which makes it more robust to adapt the unmodeled error in real cases (e.g. friction and amount of US gel). The performance of the model-based method is better than that of the local minima based method [11] on a gel-wax phantom ($2.5 \pm 1.3°$ vs $3.1 \pm 1.0°$). Furthermore, the lowest error reported in [11] is $1.7°$ while the best of the out-of-plane orientation achieved by the present work has an error of zero (see Fig. 2.23 (a)). Based on experiments performed on a phantom, the proposed method outperforms other existing methods, and is comparable to the participating human operators in terms of identification accuracy.

In-vivo validation

To further investigate the performance of the proposed probe orientation adjustment method concerning the real clinical applications, validations were carried out on volunteer's tissues (forearm, upper arm, leg, and back) using the linear probe. To limit the negative effect caused by breathing, the volunteers are asked to hold their breathing for the experiments on their low back. Ten rotations are executed for each tissue, starting from a random location. The absolute angular error for the in-plane and out-of-plane parts are shown in Fig. 2.24.

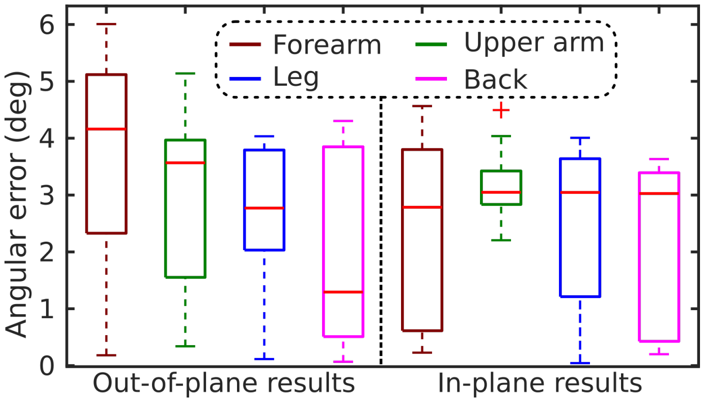


Fig. 2.24. Absolute errors for the estimation of the desired orientation $\mathbf{N}_d$ on in-vivo tissues. *Reprint from [9] with permission of IEEE.*

The error corresponding to the various in-vivo tissues ranges from $0.1°$ to $6.0°$. Regarding the out-of-plane rotation, the mean errors (±SD) are $3.7 \pm 1.7°$, $3.1 \pm 1.5°$, $2.5 \pm 1.3°$, and $2.0 \pm 1.8°$ for the forearm, upper arm, leg, and lower back, respectively. This result outperforms that of the previous state-of-the-art approach that was evaluated on a volunteer's forearm, upper arm and back achieved mean (±SD) of $3.7 \pm 1.7°$, $5.3 \pm 1.3°$, $6.9 \pm 3.5°$ [11]. Regarding the in-plane part, the mean errors (±SD) for the same tissues are $2.3 \pm 1.6°$, $3.2 \pm 0.6°$, $2.5 \pm 1.4°$ and $2.3 \pm 1.4°$, respectively. The average errors for the in-vivo tissues are close to the results on the gel-wax phantom. Therefore, we confirmed that the proposed method is partly robust against the distortions caused by unexpected human movement and tissue deformation.

In addition, it is also noteworthy that the best in-plane result on the upper arm ($2.2°$) is worse than the best results on other tissues ($0.2, 0.0,$ and $0.2°$ for forearm, leg, and back, respectively). This is due to the different fat distribution, which makes the upper arm softer than the other tissues. In addition, the round structure of upper arm is more prone to yield sliding between the probe and the tissue, which led to further deterioration of the algorithm accuracy. However, despite these challenges, the worst case in-plane results on the upper arm are still less than 4.0°.

2.6 Conclusion

In this chapter, two distinct approaches are developed for automatically orienting a US probe in the normal direction of a constraint surface without any prior knowledge. These methods enable RUSS to automatically obtain US scans with higher quality. The approach developed based on the confidence map and contact force (Sec. 2.4) is only able to be used for convex probes, while the mechanical model based approach (Sec. 2.5) works both for both linear and convex probes. Both of these approaches achieve accurate estimation of the normal direction of the contact surface for on gel phantom and in-vivio tissues (forearm, upper arm, and back). Due to the accurate force sensor, the estimation accuracy of the mechanical model based approach is slightly better than the image and force based approach (in-plane: $2.7 \pm 1.9^\circ$ vs $3.1 \pm 1.0^\circ$; out-of-plane: $1.7 \pm 0.9^\circ$ vs $2.9 \pm 0.2^\circ$). In the experimental environment, the camera-based approach [56] is only $7.3 \pm 3.1^\circ$. In both methods, the friction is not taken into consideration due to the use of US gel.

Nevertheless, the limitations of the proposed method are also outlined. First, the proposed method was developed for orthopedic applications (e.g. limb) rather than the very soft tissue (e.g. breast) as the contact could severely change the original curvature of soft anatomies, and even the optimal orientation is not necessarily to be normal anymore. Second, if the normal direction of the underlying bone surface differs considerably from the tissue surface, the orientation optimization procedure had to be performed multiple times at different positions on the scan path.

To conclude, we consider that the proposed approach will contribute significantly to developing novel RUSS for automatic day-to-day clinical examination by enabling the automated adjustment of the probe orientation, thereby achieving better image quality. Future research work will also focus on camera imaging to automatically generate a dynamic scan path and investigate the optimal contact force for different tissues.

Motion-Aware Robotic Ultrasound Imaging

3

3.1 Overview and Publications

This chapter presents the contributions towards motion-aware robotic US imaging. Robotic 3D US imaging has been seen as a promising way to overcome the limitations of traditional US examination, i.e., high inter-operator variability and lack of repeatability. In practice, human sonographers react to patient movements by repositioning the probe or even restarting the acquisition. Furthermore, several attempts for adjustment of objects are often necessary to clearly and completely image the anatomy of interest, e.g., adjusting limbs to acquire images of their entire artery tree. However, attempted adjustment of objects is not allowed for conventional robotic US systems.

To address this challenge, we propose a vision-based robotic US system that can monitor the object's motion and automatically update the sweep trajectory to provide 3D compounded images of the target anatomy seamlessly. To preliminary validate the feasibility of the ideal bring a motion-aware framework towards robust and accurate 3D object geometry, **the passive marker based approach** is first proposed. In this method, a depth camera is employed to extract the manually planned sweep trajectory after which the normal direction of the object is estimated using the extracted 3D trajectory. Subsequently, to monitor the movement and further compensate for this motion to accurately follow the trajectory, the positions of firmly attached passive markers are tracked in real-time. Finally, a step-wise compounding was performed. The experiments on a gel phantom demonstrate that the system can resume a sweep when the object is not stationary during scanning.

To further explore the effective pipeline in real scenarios, **the surface registration (marker-less) based approach** is developed. The US scan trajectory is extracted from a general CT atlas, in which the target object has been segmented. The motion monitoring system is based on the real-time segmented object masks obtained from RGB images. Once the subject moves, the robotic US will stop and automatically update its trajectory by registering the surface point clouds extracted from a depth camera before and after the movement using iterative closest point algorithm. Afterward, to ensure contact condition after repositioning of US probe, a confidence-based orientation optimization is employed to fill in the potential gap between probe and contact surface. Finally, the whole system is fully validated on a human-like arm phantom with uneven surface. The results demonstrate that the system can automatically resume a sweep when the subject moves during scanning.

This chapter is organized as follows: In Sec. 3.2, the clinical and methodological background has been summarized. Afterwards, Sec. 3.3 and Sec. 3.4 present the **passive marker based**

approach validated on a gel phantom and the **marker-less based approach** evaluated on a human-like arm phantom, respectively. Finally, the summary is presented in Sec. 3.5.

Substantial parts of this chapter have already been published and quoted verbatim. The methodology, implementation details, and results of the **the passive marker based approach** have been published in the following article.

[6] **Zhongliang Jiang***, Hanyu Wang*, Zhenyu Li, Matthias Grimm, Mingchuan Zhou, Ulrich Eck, Sandra V Brecht, Tim C Lueth, Thomas Wendler, Nassir Navab. "Motion-Aware Robotic 3D Ultrasound," In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 12494-12500, 2021 (*equal contribution).

Copyright Statement. ©2021 IEEE. With kind permission of IEEE.

The related contents of the **marker-less based approach** have been published in the following article.

[3] **Zhongliang Jiang***, Nehil Danis*, Yuan Bi, Mingchuan Zhou, Markus Kroenke, Thomas Wendler, Nassir Navab. "Precise Repositioning of Robotic Ultrasound: Improving Registration-Based Motion Compensation Using Ultrasound Confidence Optimization," *IEEE Transactions on Instrumentation and Measurement*, Early access, 2022.

Copyright Statement. ©2022 IEEE. With kind permission of IEEE.

3.2 Introduction

3.2.1 Clinical Motivation

The Peripheral Artery Disease (PAD) is one of the most common diseases, particularly among the elderly, which causes the blocking or narrowing of peripheral blood vessels, thereby limiting blood supply to certain body parts. PAD usually means a narrowing of peripheral arteries, which will further result in reduced blood flow supply to limbs [90]. The most common type of PAD is atherosclerosis, which is caused by the buildup of fats and cholesterol in and on artery walls (plaque). The illustration of atherosclerosis has been given in Fig. 3.1. Since the leg is often under a certain pressure, PAD often affects the limb arteries, particularly for the low limb arteries. The clinical complications include an infection or tissue death. In the worst case, PAD causes organ and limb failure, potentially resulting in long-term damage or death. It has been reported that PAD affects 20% of people older than 55 years, which is estimated to be about 27 million people in Europe and North America [91]. Besides, Fowkes *et* *al.* also reported that PAD affected over 200 million in 2010 globally and had become a major cause of morbidity and mortality [92].

In order to maximally restraint the worst impact of the PAD, it is very crucial to identify the PAD in early stage. In the current clinical practice, the most common solution to diagnose PAD is either by measuring the Ankle-Brachial Index (ABI) or by inspecting a Computed Tomography Angiography (CTA). The drawback of CTA is that the scans expose both the patient and medical staff to ionizing radiation. Conversely, ABI despite being radiation-free, is unable to provide the location of the PVD and is highly user-dependent [93]. Considering the aforementioned limitations of ABI and CTA, US imaging is employed as the primary non-invasive modality in clinical practice [94], because it is non-invasive, cheap, and radiation-free. Besides, Collins *et al.* reported that the femoral artery US has $80\% - -98\%$ sensitivity in detecting vessel stenoses, even for PAD in its early stages [95]. Therefore, regular US examination is important for patients.

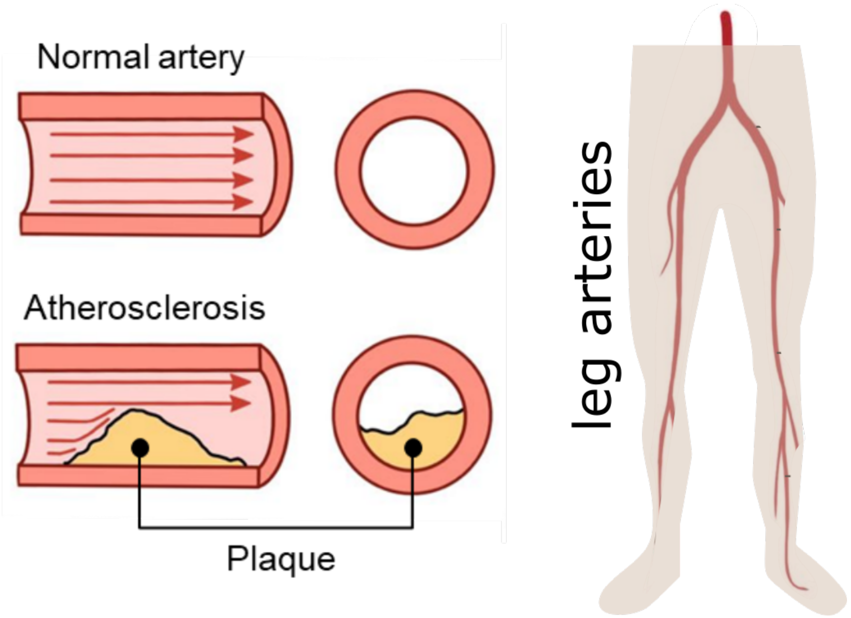


Fig. 3.1. Illustration of PAD on lower limb arteries. The Comparison between normal and abnormal arteries is depicted in the left part, and the leg arteries tree is shown in the right part.

3.2.2 Robotic 3D US Imaging

Nevertheless, due to the inherited limitations of the US imaging, conventional 2D B-mode images suffer from high intra- and inter-operator variability. These variations result in inconsistent diagnosis results of examinations carried by different sonographers, or even by the same operator at different times. To address this challenge, 3D US was employed to characterize and quantify stenoses non-invasively for carotid [96] and lower limb arteries [63]. Compared with 2D images, 3D volume is more repeatable for human operators because it does not require exactly repeating the probe orientation.

To augment 2D images into 3D volumes, Hossack *et al.* designed a probe with a 2D arrangement of transducer elements instead of traditional 1D arrays [97]. This system allows to directly visualize objects of interest in 3D, while the cost of such systems is high and the imaging quality is decreased because of the rectangular arrangement of US elements. Furthermore, such devices cannot depict structures significantly larger than their size. Besides, an optical or electromagnetic tracking system is often used to reconstruct a volume from tracked 2D images [98]. However, potential magnetic interference and occlusion between the optical camera and the markers limit the usability of such systems in real scenarios. To avoid these problems, Prevost *et al.* recorded tracking information using an inertial measurement unit and further computed the 3D artery geometry for the whole leg artery tree [99]. Nevertheless, due to the inconsistent contact force between probe and object during the long sweep, the non-homogeneous deformation in 2D images impairs the accuracy of the reconstructed 3D anatomy geometry [5].

To further address these challenges, robotic techniques are employed to develop RUSS for accurate and repeatable US acquisitions. To alleviate the influence caused by non-homogeneous deformation, two thin force sensors were attached on the tip of a US probe to maintain a constant force during the sweep [56, 57]. Gilbertson *et al.* developed an impedance controller for an one DOF device to stabilize US imaging during scans [42]. Besides contact force, Jiang *et al.* quantitatively optimize the orientation using both US imaging and estimated force at the end-effector to improve the imaging quality [11]. Additionally, being benefited from the integrated force sensors, Sun *et al.* [100] and Jiang *et al.* [5] proposed approaches to further recover the force-induced deformation in 2D and 3D images, respectively.

Since the probe poses can be computed using the known kinematic model, RUSS was also employed to compound 3D US volumes. Huang *et al.* produced 3D US volumes using a linear stage [57], and subsequently extended the system to a 6-DOF (degrees of freedom) robotic arm [56]. Virga *et al.* computed a scan trajectory from a preoperative MR image and further transferred it into the robotic frame by registering the MR to the object surface captured using an RGB-D camera [53]. Nevertheless, potential movements of patients during the scan were not taken into consideration, which may lead to failure of visualizing target anatomies. Considering tasks to display long structures, e.g., a limb artery tree, sonographers even need to actively adjust the limb pose to overcome the limitation of robotic working space and to better visualize the target artery. If the potential motion is not taken into consideration during scans, the imaging quality is significantly decayed in presence of motion as Fig. 3.2 on a gel phantom with a flat surface and a human-like arm phantom with an uneven surface.

3.2.3 Contributions

In this chapter, we first consider the potential object motion during RUSS scans. To achieve accurate and complete 3D US volume in presence of object motions during scanning, it is important to develop the approaches allow repositioning of the probe. To enable the motion-aware ability, an RGB-D camera is employed to monitor and compensate for potential object movements. This technique is particularly important for the development of robotic screening of PAD on limb arteries, where sonographers usually need to actively adjust the limb position to avoid exceeding working space limits and to visualize the whole artery tree. In this manner, RUSS can combine the advantages of free-hand US (flexibility) and RUSS (accuracy and stability).

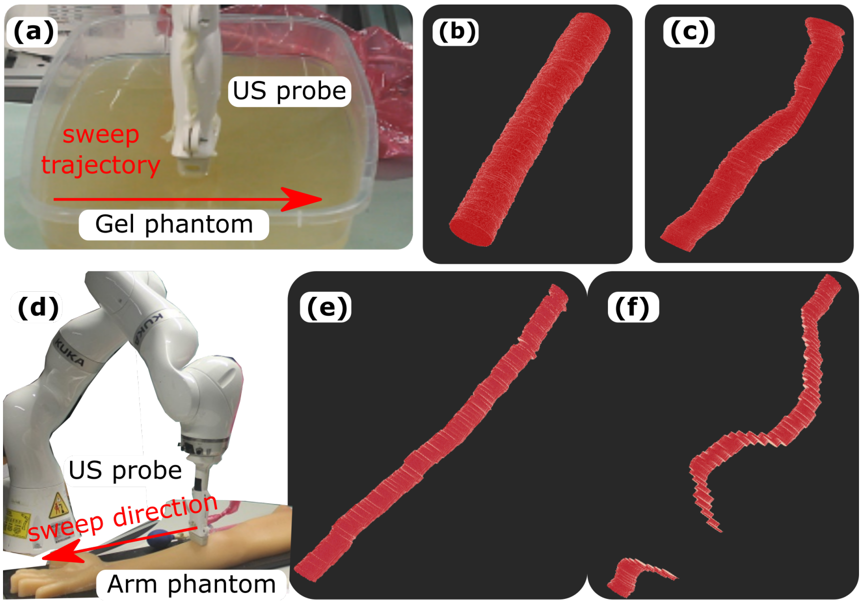


Fig. 3.2. Illustration of the influence of object motion on 3D US compounding with a gel phantom and a human-like arm phantom. (a) and (d) are US sweep on a gel phantom containing a straight tube and a uneven arm phantom with a curve artery inside. (b) and (e) 3D reconstruction result of an underlying vessel when the imaged object is stationary during the sweep, and (c) and (f) 3D result of the same vessels obtained when the object is moved randomly relative to the initial trajectory during the sweep.

To validate the feasibility of the idea of motion-aware RUSS, the **passive marker based approach** is first proposed. In this work, we present a vision-based, semi-automatic RUSS to generate complete and accurate 3D US volumes of non-stationary imaged objects during US sweeps. The main contributions can be summarized as follows:

- An adaptive method is proposed to robustly extract planned sweep trajectories drawn on the object's surface and automatically compute key points of trajectories to optimize probe poses for automatic US sweep.
- A vision-based movement-monitoring system is proposed to monitor the potential movement of the imaged object and automatically resume the sweep at breakpoints to provide a complete and accurate 3D US volume using a stepwise compounding.

The performance of the proposed motion-aware RUSS is validated on a custom-designed vascular phantom.

In addition, To further explore the effective pipeline in real scenarios, the **markerless approach** is developed. In this work, we proposed an marker-less motion-aware RUSS with the ability to reposition the probe for precise and complete 3D images of target anatomies in presence of rigid motions during scanning. Compared with the marker-based approach, a markerless approach is more convenient for operators without the requirement for carefully configuring the markers for individual patients. The main contributions are summarized as follows:

- A confidence-based approach is used to adjust the probe in-plane orientation to overcome the suboptimal contact condition between probe and object after repositioning the probe to the breaking points. Besides the aim of guaranteeing the image quality, this also benefits the visibility of the target object in the imaging view.

- A markerless vision system is developed to monitor the object's movement and automatically update the initial trajectory using the Iterative Closest Point (ICP) algorithm. The two point clouds of object surface (before and after the motion) acquired from a depth camera are registered to estimate the rigid transformation.

Finally, the segmentation approach are validated on both human arms and an arm phantom. In addition, the proposed motion-aware RUSS is fully validated on the human-like arm phantom with an uneven surface.

3.3 Marker-based Motion-Aware System

3.3.1 Overview of the Vision-Based 3D US System

The limb arterial tree is much longer than other organs, such as liver and kidney. Due to the limitation of the robotic workspace, conventional RUSS cannot handle the case when the desired trajectory is partly out of the robotic workspace. To address this challenge, we propose a RUSS allowing active adjustment of the object during the sweep to completely display the anatomy with long geometry. The overall workflow is depicted in Fig. 3.3.

Hardware

The RUSS comprises a robotic manipulator (LBR iiwa 14 R820, KUKA GmbH, Germany) and a Cicada LX US machine (Cephasonics, USA). A CPLA12875 linear probe (Cephasonics, USA) is attached to the end-effector of the robotic manipulator and B-mode US images (50 fps) are obtained via a USB interface. The robot is controlled via the iiwa stack developed based on ROS [48]. The control commands and the robot status are exchanged at $100\ Hz$. In addition, an RGB-D camera (Azure Kinect, Microsoft Corporation, USA) is used to provide real-time RGB-D images.

Proposed Workflow

The software system consists of three parts: 1) a vision-based sweep trajectory extraction, 2) an automatic robotic US sweep and 3D US compounding, and 3) a movement-monitoring system, which updates the trajectory and corrects the compounding result. To this end, an RGB-D camera and passive markers (NDI, Ontario, Canada) are used to obtain three inputs (an RGB image, a depth image, and the positions of the markers). Before performing the sweep, the desired trajectory providing good visibility of the target anatomy is manually given on the patient's skin by medical experts. The trajectory is drawn using a red color due to the strong contrast to the skin color. Afterward, the sweep extraction module is used to extract the drawn trajectory and transform it into the robotic base frame using the hand-eye calibration results. To avoid non-homogeneous deformation and guarantee the patient's

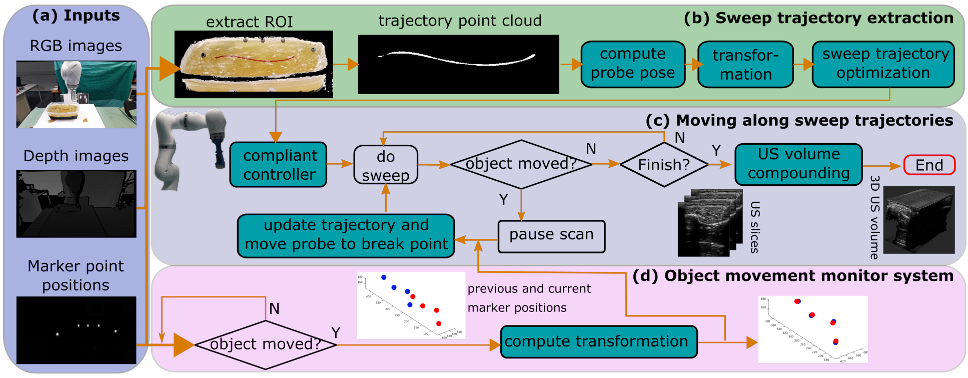


Fig. 3.3. System workflow. (a) Three types of inputs from the depth camera, (b) sweep trajectory extraction module, (c) robotic movement module following a planned trajectory, and (d) object movement monitor module. *Reprint from [6] with permission of IEEE.*

safety, a compliant controller maintains a constant contact force in the probe centerline [9]. During scanning, a marker-based motion monitor system is activated so that the system can automatically compute the transformation and update the trajectory using the ICP method if a motion happens. This correction enables a stepwise compounding.

3.3.2 Hand-Eye Calibration

To enable the ability to follow the manually drawn trajectory on the object, the transformation matrix between the RGB-D camera and the robot was calculated. It can be seen from Fig. 3.4, the involved coordinate frames are: 1) the image frame {i}; 2) the RGB-D camera frame {c}; 3) the robotic base frame {b}; 4) the robotic flange frame {f}; 5) the tool center point {tcp}; 6) an ArUco marker [101] {ar}. The transformation from the flange to the base frame ${}^{b}_{f}\mathbf{T}$ can be directly obtained using the robotic kinematic model. Thus, the transformation ${}^{f}_{tcp}\mathbf{T}$ is obtained using the 3D model of the custom-designed probe holder (connecting the US probe to the robot). Besides, the transformation ${}^{c}_{i}\mathbf{T}$, used to generate a 3D point cloud from the 2D image can be computed based on the camera intrinsics obtained via a program[1].

The transformation between frame {b} and {c} ${}^{b}_{c}\mathbf{T}$ can be optimized based on the paired descriptions of the points in different frames using ICP [16]. To accurately estimate ${}^{b}_{c}\mathbf{T}$, at least four non-coplanar points should be employed. Thus, eight arbitrary intersections on two chessboards at different heights are selected. For each point, the position in the camera frame ${}^{c}P$ is computed using OpenCV and the coordinate expression in frame {b} ${}^{b}P$ is obtained by manually moving the robotic arm to the target intersection. To accurately locate the intersections, a tip-pointed tool, was attached to the flange. Based on the paired position descriptions (${}^{c}P$ and ${}^{b}P$), the transformation ${}^{b}_{c}\mathbf{T}$ is obtained by optimizing the following equation:

$$\min_{{}^{b}_{c}\mathbf{T}} \frac{1}{N} \sum_{i=1}^{N} ||{}^{b}_{c}\mathbf{T}\, {}^{c}P_i -{}^{b} P_i||^2 \tag{3.1}$$

---

[1] https://github.com/microsoft/Azure_Kinect_ROS_Driver

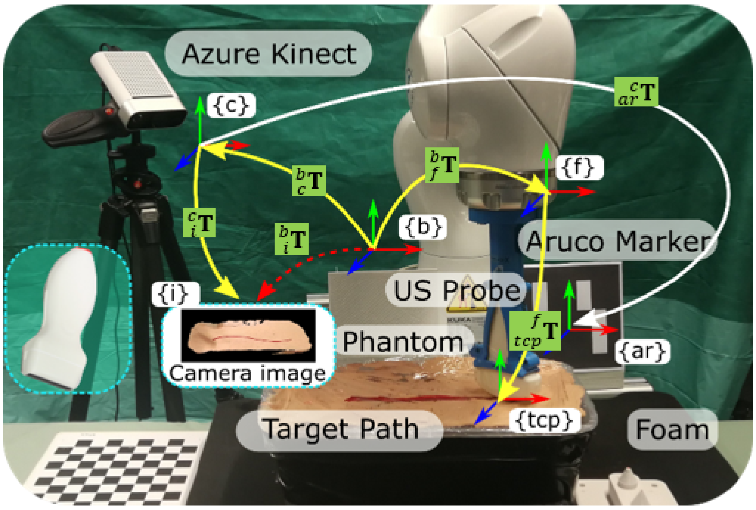


Fig. 3.4. Diagram of the involved coordinate frames. *Reprint from [6] with permission of IEEE.*

To make the calibration system robust to the camera movement, an ArUco marker is fixed relatively to the robotic base as shown in Fig. 3.4. Once ${}^{b}_{c}\mathbf{T}$ is calculated using Eq. (3.1), the fixed ${}^{b}_{ar}\mathbf{T}$ can be calculated using ${}^{b}_{ar}\mathbf{T} = {}^{b}_{c}\mathbf{T} \cdot {}^{c}_{ar}\mathbf{T}$. Then, based on real-time ${}^{c}_{ar}\mathbf{T}$, obtained using aruco_ros[2], ${}^{b}_{c}\mathbf{T}$ can be dynamically updated when the camera is moved.

3.3.3 Extraction of Scan Trajectory

In this section, we describe the adaptive color-based method for extracting the manually planned trajectory on object's surface. To provide reliable information on the object's position, two passive markers are fixed on the skin (using tape) in both ends of the trajectory as Fig. 3.3. The marker spheres are covered by a retro-reflective layer, making the markers brighter than the background in the infrared images. Also, the markers aid in robust extraction of the region of interest (ROI) and trajectory (whole trajectory or at least partial trajectory if it is partly occluded by surrounding stuffs).

A multi-color space threshold method was used to exact the ROI [102]. The ROI is extracted based on the color feature in the area surrounding the two markers at the end of the trajectory in the Cb and Cr channel images. The upper and lower limits of the ROI are automatically determined based on the value of pixels distributed on the line connecting the two markers. Compared with RGB images, YCrCb images separate brightness and chromaticity into different channels. Thus, the negative effect caused by the environment (e.g. light) could be limited using the YCrCb color space. To further extract the red trajectory from the ROI, an adaptive threshold method on both the Cr channel image and the grayscale image is proposed. To robustly extract the trajectory line, even when partly blocked by the probe, N_s seeds lines are initialized to equally divide the space between the two markers at the end of the trajectory. Then, the intersection points P_s between the seed line and the trajectory can be detected by locating the maximum Cr value of the pixels on the seed line (Fig. 3.5).

---

[2] https://github.com/pal-robotics/aruco_ros

However, since the US probe may partly block the trajectory in RGB images, the detected points P_s may not be on the trajectory. To further remove these points from P_s, the "up" (Y_p) and "down" (-Y_p) boundaries of the trajectory for each P_s in the Cr channel images can be calculated using:

$$\begin{aligned} f_b^{up} &= max\left([|I(x_i, y_i + j + 1) - I(x_i, y_i + j)|]\right) \\ f_b^{down} &= max\left([|I(x_i, y_i - j - 1) - I(x_i, y_i - j)|]\right) \end{aligned} \tag{3.2}$$

where (x_i, y_i) represents the i-th P_s, j is $1, 2, 3, ..., 8$.

If the boundary features f_b^{up} and f_b^{down} are close to zero (< 10), the boundary features on the grayscale images are computed again using Eq. (3.2). If the features for $P_s(i)$ computed based on the grayscale images are still close to zero, it will be removed from P_s because the intensity of the real P_s are supposed to be significantly different from the skin background.

Based on the computed seed points P_s located at different positions of the trajectory, a bidirectional searching method starting from each P_s is developed. Considering the potential color differences between different trajectory parts, an adaptive threshold is used to effectively extract the trajectory. First, a moving box ($B_w \times B_h$) is initialized at $P_s(i)$. Then, the number of points, whose intensity is between $[I(P_s))$-$I_rt,\ I(P_s) + I_rt]$, is counted as N. If N is larger than the empirical threshold N_{ep}, at least a short part of the trajectory is located inside the current box. To further extract the trajectory, the moving box and the corresponding local threshold range are updated using the last detected trajectory point. These procedures have been described in lines $5 - 18$ in Algorithm 1. Similarly, another moving box is initialized at $P_s(i)$ to search in the inverse direction. The seeds-based bidirectional searching modality enables the proposed method to provide the most possible result. The involved parameters are empirically set to: $B_w = 20\, pixels$, $B_h = 100\, pixels$, $I_{rt} = 25$, and $N_{ep} = 10$.

3.3.4 Probe Orientation Determination and Trajectory Optimization

Based on the 2D trajectory P_t^{2d} extracted in the last section, the 3D trajectory point cloud P_t^{3d} is computed using the camera internal parameters and used to control the position of the probe. However, P_t^{3d} only tells the desired positions of probe during the scan. To receive more signal back to the US elements built in US probe tip, Jiang *et al.* and Ihnatsenka *et al.* suggested that the probe should be aligned in the normal direction of the contact surface $\vec{n}_i$ [11, 55]. Here, a vision-based normal direction estimation method was proposed to quickly compute the desired poses for the whole sweep rather than a force-based method, which only works for the current contact position used in [11]. Compared with [57], more local points distributed around the point on the trajectory P_t^{3d} are considered to accurately and stably estimate $\vec{n}_i$. The selected point will be located on a plane when the local points are distributed close enough. In that case, $\vec{n}_i$ at $P_t^{3d}(i)$ is approximated by the normal direction of the plane. The plane expression ($z =$ f(x, y)) is optimized using the Least-Squares method as Eq. (3.3).

Algorithm 1: Adaptive Trajectory Extraction

```
1  Input: seed points set $P_s = (x_s, y_s)$, moving box width $B_w$, moving box height $B_h$,
          relative threshold of Cr channel $I_{rt}$ and the number of the pixels extracted in the
          moving box $N_{ep}$
   Output: 2D trajectory $P_t^{2d}$
2  for $i = 2; i < len(P_s); i++$ do
3      Searching from $P_s(i)$ to $P_s(i+1)$;
4      for $x = x_s(i); x < x_s(i+1)$ do
5          $N \leftarrow 0$ ;
6          for $m = x; m < x + B_w; m++$ do
7              for $n = y_s(i) - \frac{1}{2}B_h; n < y_s(i) + \frac{1}{2}B_h; n++$ do
8                  if $I(m,n) \in [I(x, y_s(i)) - I_r t,\ I(x, y_s(i)) + I_r t]$ then
9                      $P_t^{2d} \leftarrow [P_t^{2d}, (m,n)]$ ;
10                     $N \leftarrow N + 1$ ;
11                 end
12             end
13         end
14         if $N > N_{ep}$ then
15             $(x, y_s(i)) \leftarrow (P_t^{2d}(end))$;
16         end
17         else
18             Break;
19         end
20     end
21     Similar to 3-19, searching from $P_s(i)$ to $P_s(i-1)$;
22 end
```

$$\min_{f(x_i, y_i)} \frac{1}{2N} \sum_{i=1}^{N} (f(x_i, y_i) - z_i)^2 \tag{3.3}$$

Then the normal direction of tissue surface $\vec{n}_i$ at $\mathbf{P}_i$ is approximated by the normal diction of the optimized plane. Also, the radius of the sphere is also crucial for the estimation performance. On the one hand, a large radius will cause large bias in estimation of normal direction of a point. On the other hand, a small radius means that the estimation result is not stable and it is easy to be affected by the noise measurements. In this work, the radius is empirically selected as 1 cm.

After aligning the probe to the estimated $\vec{n}_i$, the probe tip is expected to be perpendicular to the scan trajectory. Since the width of the manually drawn trajectory varies, the moving direction computed by connecting two close points in P_t^{3d} may differ significantly from the real value, causing instability in the rotation around the probe centerline (aligned with $\vec{n}_i$) during scanning. To address this problem, we propose a difference-based optimization method to automatically select key points from P_t^{3d}, generating a smooth robotic movement trajectory. To this end, the 3D points P_t^{3d} are transformed into 2D vectors $(x_p(i), y_p(i))$ as follows:

$$(x_p(i), y_p(i)) = (\overrightarrow{P_s P_t^{3d}(i)} \cdot \mathbf{X}_p,\ |\overrightarrow{P_s P_t^{3d}(i)} \times \mathbf{X}_p|) \tag{3.4}$$

where $\mathbf{P}_s$ is the start point of the path and $\mathbf{X}_p$ is a unit vector connecting start and end points $\mathbf{P}_e$ as $\mathbf{X}_p = \frac{\mathbf{P}_e - \mathbf{P}_s}{|\mathbf{P}_e - \mathbf{P}_s|}$.

Based on the transformed 2D position vector $P_p(x_p, y_p)$, the local maxima and minima are extracted as follows:

$$P_k(i) \;\; i \in \{k | D(k-1)D(k) \leq 0 \& |D(k-1)| + |D(k)| > T_k\} \tag{3.5}$$

where $D(i) = y_p(i+1) - y_p(i)$ is the first-order difference and T_k is the threshold used to remove the local extrema with a small amplitude. Fig. 3.5 shows an illustration of the key points in both 2D and 3D, where all turning points have been correctly detected as key points (marked as red circles).

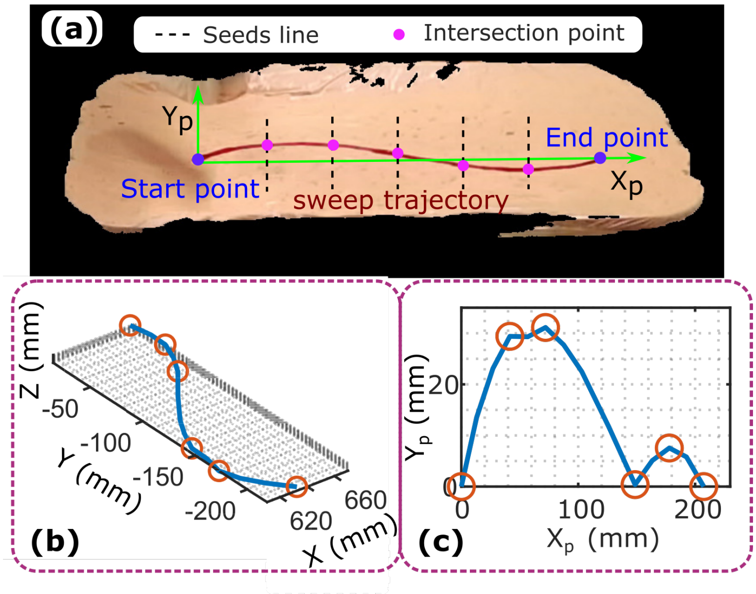


Fig. 3.5. Trajectory optimization of the robotic movement. (a) Planned trajectory on the object's surface. (b) and (c) are automatically detected key points in 3D space and 2D space, respectively. *Reprint from [6] with permission of IEEE.*

To restrain the effect caused by the local identification error of estimated probe orientation, the orientation is further optimized for each interval separated by the key points and the start and stop points. To improve movement stability, points located close to their neighbors are not used to define an interval as lines $5-7$. For each interval, the desired direction of the probe centerline (${}^{tcp}\mathbf{Z}$) is represented by the mean $\vec{n}_i$ computed at all sampled positions. Afterward, the probe's long side direction (${}^{tcp}\mathbf{Y}$) is placed in the normal direction of the plane consisted by ${}^{tcp}\mathbf{Z}$ and the line connecting the two $\mathbf{P}_k$ defines the interval. Finally, the moving direction (${}^{tcp}\mathbf{X}$) is perpendicular to ${}^{tcp}\mathbf{Z}$ and ${}^{tcp}\mathbf{Y}$. The implementation details are described in Algorithm 2.

3.3.5 Movement Monitor System

To monitor potential object motions, the markers attached to the ends of the trajectory are used as inputs to the monitor system. The robot immediately stops when the system detects a

Algorithm 2: Key points Based Orientation Optimization

```
Input: Planned path $\mathbf{P}_t^{3d}$, start point $\mathbf{P}_s$, end point $\mathbf{P}_e$, Key points on path $\mathbf{P}_k$, normal
       direction of tissue $\vec{n}_i$ at each $\mathbf{P}_t^{3d}(i)$, and threshold $T_1$
Output: Segmented intervals Segs, optimized probe orientation of all intervals
        $[^{tcp}\mathbf{X}, ^{tcp}\mathbf{Y}, ^{tcp}\mathbf{Z}]$
Add $\mathbf{P}_s$ and $\mathbf{P}_e$ to $\mathbf{P}_k$: $\mathbf{P}_k \leftarrow [\mathbf{P}_s, \mathbf{P}_k, \mathbf{P}_e]$ ;
Compute iteration set of $\mathbf{P}_k$ in $\mathbf{P}_t^{3d}$: $\mathbf{I} = find(\mathbf{P}_k)$;
int $m = 1$;
for $j = 1; j < len(\boldsymbol{P}_k)$ do
    if $|\boldsymbol{I}(j+m) - \boldsymbol{I}(j)| * |\boldsymbol{P}_k(j+m) - \boldsymbol{P}_k(j)| \leq T_1$ then
        $m \leftarrow m + 1$;
    end
    else
        Segs $\leftarrow$ [Segs, $[I(j), I(j+m)]$];
        $^{tcp}\mathbf{Z} \leftarrow [^{tcp}\mathbf{Z}, \frac{1}{I(j+m)-I(j)} \sum_{i=I(j)}^{I(j+m-1)} \vec{n}_i]$;
        $^{tcp}\mathbf{Y} \leftarrow [^{tcp}\mathbf{Y}, \frac{1}{I(j+m)-I(j)} \sum_{i=I(j)}^{I(j+m-1)} \vec{n}_i \times \overrightarrow{\mathbf{P}_t^{3d}(I(j))\mathbf{P}_t^{3d}(I(j+m))}]$;
        Pose = [Pose, $[^{tcp}\mathbf{Y} \times ^{tcp}\mathbf{Z}, ^{tcp}\mathbf{Y}, ^{tcp}\mathbf{Z}]$];
        $j \leftarrow j + m$;
        $m \leftarrow 1$;
    end
end
$[^{tcp}\mathbf{X}, ^{tcp}\mathbf{Y}, ^{tcp}\mathbf{Z}] \leftarrow$ Pose;
```

change in the position of marker over a given threshold ($5\ mm$). To accurately update the trajectory to resume the sweep from the break point, three additional markers are randomly stitched on the object. Next, the transformation between the previous and current object poses is calculated using ICP based on four paired markers positions as shown in Fig. 3.3 (d). The last marker is used to compute the error of the ICP results as follows:

$$e_{mc} = \left\| \mathbf{P}_m^{'} - (\mathbf{R}_{mc} \cdot \mathbf{P}_m + \mathbf{T}_{mc}) \right\| \tag{3.6}$$

where $\mathbf{P}_m$ and $\mathbf{P}_m^{'}$ are the positions of the passive markers before and after the movement, respectively, while $\mathbf{R}_{mc}$ and $\mathbf{T}_{mc}$ are the computed rotation matrix and translation vector.

The RUSS automatically resumes the sweep from the breakpoint only when e_{mc} is small enough ($< 1\ cm$). Otherwise, the RUSS automatically stops the sweep.

3.3.6 Results

Vision-based Trajectory Extraction Results

To validate the performance of the adaptive trajectory extraction method described in Sec. 3.3.3, a gel phantom with a manually drawn trajectory on its upper surface was employed. The phantom was randomly placed in different places inside the camera view. To demonstrate the result, an RGB image with the detected trajectory is presented in Fig. 3.6 (a) and (c), respectively. Besides, considering that the trajectory will be partly occluded by the robotic arm and the US probe during the sweep, the probe on the top of the trajectory was positioned as in Fig. 3.6 (b). The proposed adaptive method completely displayed the unblocked part of the trajectory as Fig. 3.6 (d). The detected results are close to the trajectory in Fig. 3.5 (a). More examples of the detected results can be found in this video[3].

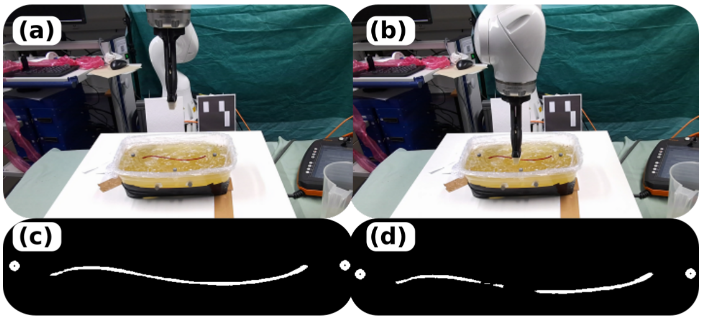


Fig. 3.6. Results of the trajectory extraction method. RGB images obtained from the depth camera when the trajectory can be (a) completely or (b) partially seen, depending on the probe position. (c) and (d) are the detected results of (a) and (b), respectively. The line represents the detected trajectory, while the white circles represent the passive markers. *Reprint from [6] with permission of IEEE.*

Since the direction of the probe centerline (${}^{tcp}\mathbf{Z}$) is controlled in the force mode [9], the probe tip along the trajectory in the ${}^{tcp}X$–${}^{tcp}Y$ plane should be accurately moved. To quantitatively analyze the whole system accuracy (trajectory detection and hand-eye calibration), the computed trajectory was transformed into the robotic frame using the result of hand-eye calibration. To demonstrate the real position error between the desired trajectory and the performed path, the position of the probe tip was recorded. Besides, the computed error is calculated based on the computed trajectory and ground truth. The ground truth is obtained by manually guiding the robot along the manually drawn trajectory. The ground truth, the computed trajectory, and real trajectory are depicted in Fig. 3.7.

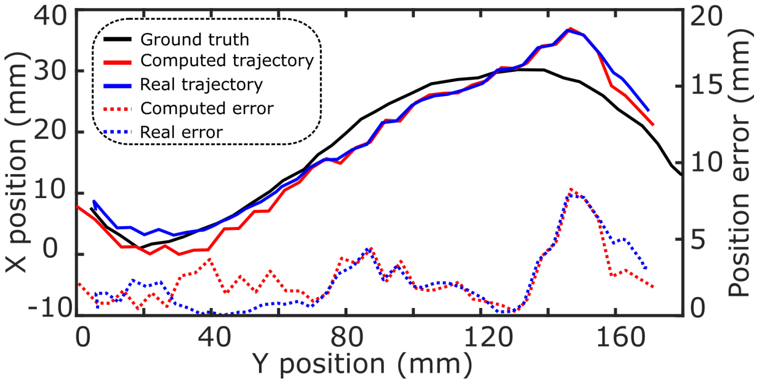


Fig. 3.7. Trajectory following results. The solid lines represent the trajectories. The dotted lines are the computed and real position error. *Reprint from [6] with permission of IEEE.*

As shown in Fig. 3.7, both the computed and real error mainly distribute below $5\ mm$. Also, the error is not cumulative. The error still can be close to zero after a large error, for example when $Y = 150\ mm$ (Fig. 3.7). Generally, the average computed and average real errors ($\pm SD$) were $2.5 \pm 1.8\ mm$ and $2.0 \pm 1.9\ mm$, respectively. The computed error was mainly caused by the depth estimation. Since our application must accurately control the probe in the ${}^{tcp}X$–${}^{tcp}Y$ plane, the accuracy of the system can be further improved by fixing the camera

---

[3] https://https://www.youtube.com/watch?v=8IaorIl3zzk

on the top of the target objects to reduce the negative influence caused by inaccurate depth estimations. In addition, the real error is smaller than the computed error. This is because the real movement of the RUSS system lags behind the control command determined based on the computed trajectory.

Movement Compensation Results

To validate the performance of the movement compensation algorithm, experiments were carried out on a gel phantom. To further investigate the method's sensitivity to the movement types (translation and rotation), the experiments were categorized into two sets of translation and rotation, respectively. For the translation set, the phantom was moved along a straight line at different distances ($50, 100, 150, \ and \ 200 \ mm$). For the rotation set, the phantom was rotated around a fixed axis at different angles ($10, 20, 30, \ and \ 40°$). The illustration of the setup is given in Fig. 3.8. Each set consisted of 10 independent experiments. The translated and the rotated axes are randomly set in the camera view.

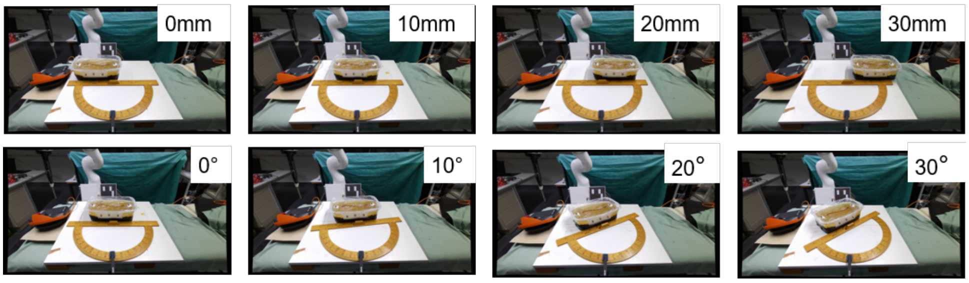


Fig. 3.8. Illustration of the experimental setup used for validate the compensation performance.

To assess the performance of the marker-based compensation method, the errors e_{mc} (Eq. (3.6)) of the two groups of experiments are presented in Fig. 3.9. The absolute translation and absolute rotation errors (±SD) were $3.1 \pm 1.0 \ mm$ and $2.9 \pm 0.7 \ mm$, respectively. Also, based on a t-test (probability $p = 0.67 > 0.05$), it was concluded that no significant difference exists between the two experimental sets. Furthermore, the results also show that most errors ($> 75\%$) are less than $4.0 \ mm$ even when the phantom was moved $200 \ mm$ and rotated $40°$. Such e_{mc} is promising in clearly displaying the partitioned sweeps, which generate complete geometry of the anatomy after a larger motion because $e_{mc} = 4 \ mm$ is much smaller than the probe width $37.5 \ mm$.

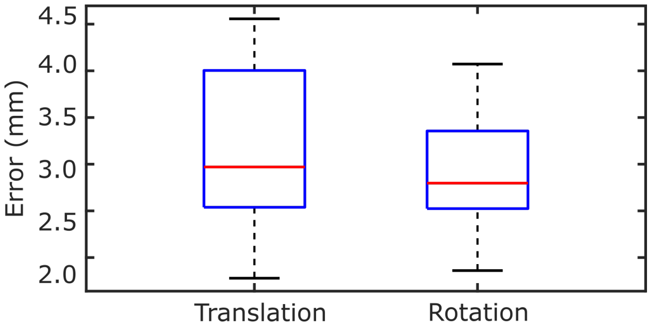


Fig. 3.9. Absolute error of the movement compensation algorithm. Two motion types (translation and rotation) over 10 trials are displayed. *Reprint from [6] with permission of IEEE.*

To intuitively show the performance of the compensation method on 3D compounded results, experiments were carried out on a gel phantom with a straight hole, mimicking the blood vessel, which was automatically segmented from the cross-sectional US images using a well-trained U-Net from [7]. Then, the 3D compounding process was performed using ImFusion Suite (ImFusion GmbH, Munich, Germany). The 3D vessels with and without motion compensation are shown in Fig. 3.10. The result reveals that the proposed approach can deliver a complete 3D image of the target blood vessel even in the presence of object motion. Limited by the accuracy of depth estimation, the stitching is still visible using the current setup. But the stitching error could be reduced using a camera with an accurate depth estimation.

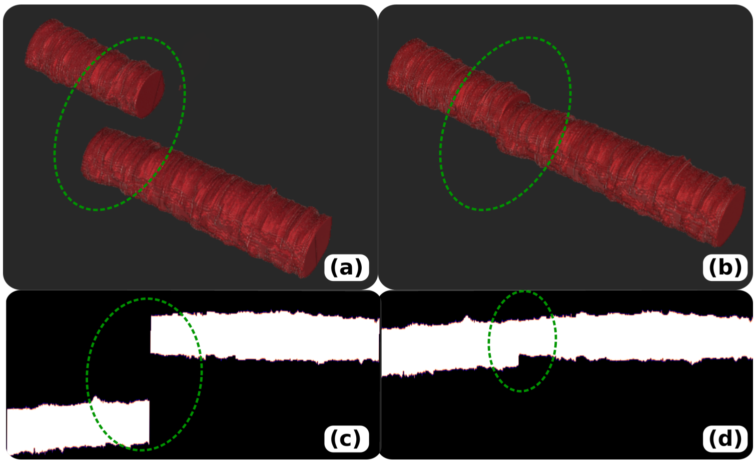


Fig. 3.10. Performance of compensation method. 3D images of a straight vessel (a) without and (b) with motion compensation. (c) and (d) are the corresponding 2D images (axial plane) of (a) and (b), respectively. *Reprint from [6] with permission of IEEE.*

3.3.7 Discussion

The preliminary validation on a gel phantom demonstrates that the proposed approach can provide a promising 3D geometry even when the scanned object is moved. In this work, the trajectory was drawn using a marker pen, which can be replaced by a laser system to automatically project a sweep trajectory on the surface. Since the relative threshold was determined based on local seed points, a promising result can be computed if the trajectory color significantly differs from the skin color. However, since the skin color was approximated using the passive markers, the visibility of the markers must be guaranteed during the initialization procedure before executing scan. This affordable system is developed to completely visualize the anatomy with long structures by automatically stitching the US sweeps acquired when the object is moved. Although the vascular application was used to demonstrate the proposed methos, the method can also be used for other applications such as US bone visualization. Regarding the motion compensation, the proposed method so far considered the large object motions during US scans. The small physiological motions ($< 5\ mm$), such as tremor, vessel motion [103] and breathing [10, 17] have not been considered. Since a compliant controller was used, the probe is flexible in its centerline direction. Thus, compared with free-hand US acquisition, the effect of such small motions is not aggravated in our setup. If we further consider the articulated limbs, more markers could be used to compute the transformation of each rigid parts.

3.4 Surface Registration based Motion-Aware System

In this section, we first describe the pipeline of the proposed surface registration based motion-aware RUSS towards an accurate and complete 3D volume of object anatomies against potential (expected or unexpected) object movements during the scanning. The overview is shown in Fig. 3.11

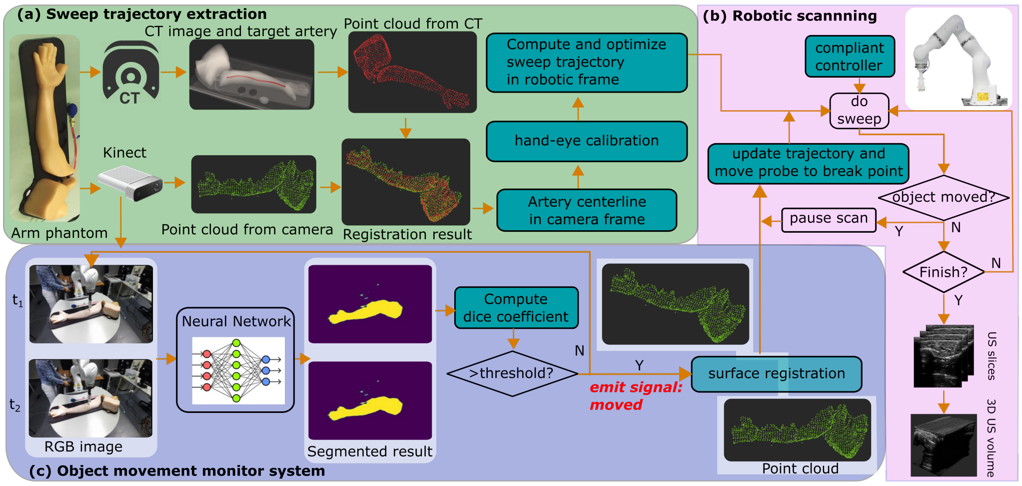


Fig. 3.11. Framework of the motion-aware RUSS. (a) the sweep trajectory extraction and optimization module, (b) the robotic scanning execution module, and (c) the camera-based movement monitoring module. The dice threshold is 0.95 in this work.

3.4.1 Overview of the Markerless 3D US system

To enable the motion-aware ability, a surface registration based approach is developed in this work to detect the (expected or unexpected) movements and update the sweep trajectory based on the surface registration approach. The general pipeline is described in Fig. 3.11. Compared with the marker-based approach [6], the passive markers are not needed anymore and a human-like arm phantom is used to mimic real scenarios.

Hardware Setup

The proposed motion-aware RUSS mainly consists of three components: a redundant robotic arm (LBR iiwa 14 R820, KUKA GmbH, Germany), a US machine (Cephasonics, USA), and an RGB-D camera (Azure Kinect, Microsoft Corporation, USA). The robot is controlled using a self-developed Robot Operating System (ROS) interface [48]. The control commands are updated at $100\ Hz$ to guarantee real-time performance. In addition, a linear probe (CPLA12875, Cephasonics) is attached to the robotic flange using a custom-designed holder. The B-mode images are accessed via a USB interface ($50\ fps$) and visualized in a software platform

(ImFusion Suite, ImFusion GmbH, Germany). The detailed configurations of US acquisition are listed as follows: imaging depth: $55\ mm$, brightness: $67\ dB$, frequency: $7.6\ MHz$. To validate the proposed system, a commercial human-like arm phantom (PICC US Training Model, Skills Med Deutschland GmbH, Germany) is employed.

Implementation Details

The motion-aware RUSS is realized by three main components: the sweep trajectory extraction and optimization module, the robotic scanning execution module, and the camera-based movement monitoring module. Regarding the extraction of sweep trajectory, the artery of interest is manually segmented from preoperative CT data using an open-source software 3D slicer[4]. Then, the meshes of the arm surface and the target limb artery surface are generated and further used to create point clouds using Meshlab[5] [see Fig. 3.11 (a)]. To transfer the target artery position from CT data to the robotic base frame, a self-occlusion point cloud of the arm surface is generated based on the RGB-D camera placed in front of the target object. By registering the CT point cloud to the camera point cloud, the artery trajectory can be transferred into the camera frame. Further utilizing the hand-eye calibration result, the artery trajectory can be transferred into the robotic base frame to guide the probe during the scanning.

Afterward, the RUSS starts executing US examination along the target artery. To provide good imaging quality and guarantee patient safety during the scanning, a compliant controller is employed to maintain a constant force between the probe and the contact surface [7, 48]. In addition, to monitor the movement of the objects, a neural network (UNet-VGG16 [104]) was trained to segment the arm from RGB images. The segmented results of images acquired at t_1 and t_2 are further used to compute the dice coefficient. Once the dice coefficient is smaller than a preset threshold, the system considers the object has moved and a corresponding signal is emitted to ROS master. After detecting such a signal, the robot stops and perform the registration between the camera point clouds acquired at t_1 and t_2 is carried out. Based on the registration result, the remaining trajectory can be updated and RUSS can continue the scan by automatically move the probe to the breaking point, namely, the point the US probe was when the movement happened. Benefiting from this correction, a 3D compounding is achieved by properly stacking the B-mode images in 3D space.

3.4.2 Coordinate System Transformation

There are two different types of calibration procedures involved in this study: 1) US calibration and 2) hand-eye calibration. The former one is used to transfer the pixel position in B-mode images to the robotic base frame while the latter one is used to transfer the position from the camera view to the robotic base frame to guide robotic movement. The implementation details are similar to the ones described in Sec. 3.3.2. The involved coordinate systems are depicted in Fig. 3.12: 1) the robotic base frame $\{b\}$, 2) the robotic flange frame $\{f\}$, 3) the probe frame $\{p\}$, 4) US imaging frame $\{us\}$, 5) the camera frame $\{c\}$, and 6) the CT atlas frame $\{ct\}$.

---

[4] https://www.slicer.org/

[5] https://www.meshlab.net/

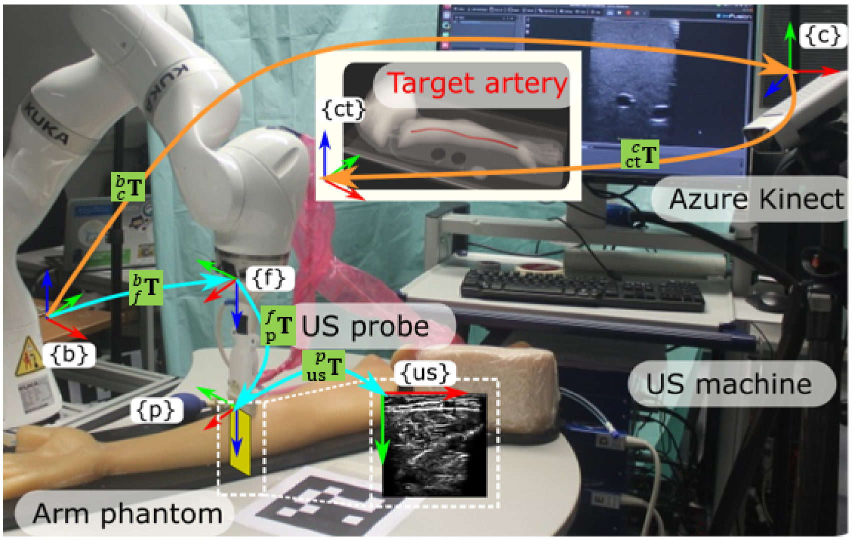


Fig. 3.12. Illustration of involved coordinate frames. Red arrows, green arrows and blue arrows represent X, Y and Z directions of corresponding coordinate systems, respectively.

To perform trajectory correction based on US imaging feedback, the transformation matrix ${}^{b}_{us}\mathbf{T}$ between frame $\{b\}$ and frame $\{us\}$ can be calculated as follows:

$$ {}^{b}_{us}\mathbf{T} = {}^{b}_{f}\mathbf{T}\,{}^{f}_{p}\mathbf{T}\,{}^{p}_{us}\mathbf{T} \tag{3.7} $$

where ${}^{j}_{i}\mathbf{T}$ is the transformation matrix used to transfer the position from frame $\{i\}$ to frame $\{j\}$. ${}^{b}_{f}\mathbf{T}$ represents the kinematic model of the robotic arm. This can be directly accessed by the API function provided by the manufacturer. Besides, ${}^{f}_{p}\mathbf{T}$ depends on custom configuration. In this work, the rotation part of ${}^{f}_{p}\mathbf{T}$ is set to an identity matrix $\mathbf{I}^{3\times 3}$ and the translational part is obtained from the custom-designed 3D holder model. Due to the characteristic of linear probe that US elements are physically distributed on the probe tip within a specific length L_p, the horizontal mapping ($\frac{L_p}{W_{us}}$) between pixel-wise positions and physical positions can be computed, where W_{us} is the pixel-wise width of B-mode imaging. Similarly, the vertical mapping is computed by $\frac{D}{H_{us}}$, where D and H_{us} are the depth setting and the pixel-wise height of US images. The origin of frame $\{us\}$ was set at the upper left of the B-mode image, while the origin of frame $\{p\}$ was set in the middle of the probe tip (Fig. 3.12). Thus, ${}^{p}_{us}\mathbf{T}$ is calculated as follows:

$$ {}^{p}_{us}\mathbf{T} = \begin{bmatrix} 0 & 0 & -1 & 0 \\ -\frac{L_p}{W_{us}} & 0 & 0 & \frac{L_p}{2} \\ 0 & \frac{D}{H_{us}} & 0 & \varepsilon_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.8} $$

where $L_p = 37.5\ mm$, $D = 55\ mm$, and hyper parameter ε_0 is used to neutralize the indeterminacy of US elements configuration.

In addition, regarding the hand-eye calibration, similar implementation procedures are carried out as [6]. To compute the transformation matrix ${}^{b}_{c}\mathbf{T}$ between the base frame $\{b\}$ and the camera frame $\{c\}$, paired coordinate descriptions of the points (intersection points on two chessboards) in frame $\{c\}$ and frame $\{b\}$ are recorded. The coordinate representations in $\{c\}$ are computed using OpenCV while the representations in $\{b\}$ are obtained by manually moving a custom-designed pointer tool to the same intersection points. More implementation details can be found in [6].

3.4.3 Surface Registration and Trajectory Transfer

Since the object position varies in different examinations, a depth camera is employed to gain information about the current environmental setup. To properly execute US scan for the target artery, the annotated artery is required to be transferred into camera frame $\{c\}$. To obtain such transformation ${}^{c}_{ct}\mathbf{T}$, the ICP algorithm is used to register the surface point cloud $\mathbf{P}_{ct}$ obtained from a pre-operative CT image to the live point cloud $\mathbf{P}_{c}$ computed from the RGB-D camera. ICP proved to be sufficient for the targeted task in preliminary experiments. To achieved uniformly distributed $\mathbf{P}_{ct}$, the Poisson disc sampling algorithm [105] is employed instead of the random sampling approach. In this work, the sizes of the arm point cloud and artery point cloud are 3000 and 1000, respectively. Since ICP is sensitive to the initial result, the multiscale feature resistance approach is used to extract unique features from both point clouds for computing the initial transformation [106]. Afterward, the iterative algorithm is used to accurately align $\mathbf{P}_{ct}$ to $\mathbf{P}_{c}$.

Since the artery is located inside the arm, the artery trajectory is required to be projected to the surface for generating an executable scanning trajectory. To alleviate potential negative influences of registration error between $\mathbf{P}_{ct}$ and $\mathbf{P}_{c}$ on creating scanning trajectory, the point cloud of the artery surface in CT $\mathbf{P}_{ct}^{art}$ is directly transferred into the camera frame. Then, $\mathbf{P}_{ct}^{art}$ can be exactly projected onto the object surface captured by the camera. Based on the hand-eye calibration result described in Sec. 3.4.2, the scanning trajectory is computed.

Artery Centerline Extraction

The artery surface point clouds have a tubular structure. Besides, regarding the geometry of limb arteries, the length along the vascular centerline is significantly longer than in other directions (namely, artery diameter). To extract the vascular centerline for generating scanning trajectory, the principal component analysis (PCA) is employed to define a new orthogonal coordinate system for optimally describing variance of the data. The first principal component is the direction making the projections achieve the largest variance, namely the eigenvector corresponding to the largest eigenvalue of the artery point cloud's covariance matrix. To estimate the vascular centerline, the average value of all points inside a small artery segmentation is computed to an approximate local center point. A small distance interval d_{in} is used to generate multiple center points between the minimal and maximal value of the projections in the first principal component. The vascular centerline is approximated by connecting the center points successively.

Scanning Trajectory on Object Surface

After deriving the vascular centerline, an executable scanning trajectory on the object surface should be further generated. To avoid potential safety issues caused by improper scanning trajectories, e.g., inside of the arm, the scanning trajectory is generated as following steps. Since the robot and the experimental table are parallel to the ground, a continuous trajectory can be created by projecting the estimated vascular centerline onto the surface in the direction of Z_b. To achieve this, the center points of the artery (x, y, z) computed in last section is replaced with (x, y, z^{arm}_{max}), where z^{arm}_{max} is the maximum value of the arm surface point cloud $\mathbf{P}_{ct}$ in Z_b direction. Then K-nearest neighbors (KNN) approach is used to search for the $K_{st} = 5$ nearest points around (x, y, z^{arm}_{max}), individually, in terms of Euclidean distance. Then, the average of these K_{st} neighbors is used as key points of the scanning trajectory for RUSS.

To fully control the probe during the scanning, the probe orientation needs to be further determined. To improve the contrast of the resulting B-mode imaging, Ihnatsenka *et al.* and Jiang *et al.* suggested that the probe should be aligned in the normal direction of the contact surface $\vec{n}_i$ [11, 55]. Here, the normal direction is quickly estimated using computer vision techniques based on the neighboured points around the key points. Compared with the force-based method [11], the vision-based method is able to quickly compute all the normal directions along the whole trajectory rather than only the current contact point. This process is implemented using Point Cloud Library (PCL)[6]. The estimated normal directions at various positions are shown in Fig. 3.13. To properly visualize the target vessel, the probe centerline is aligned to the computed normal direction and the long axis of the probe (Y_p) is aligned to be orthogonal to the scanning path.

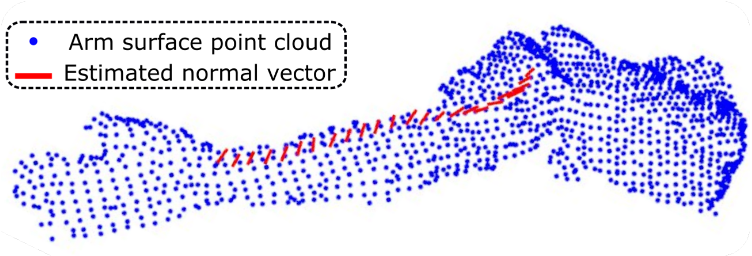


Fig. 3.13. Computed scanning trajectory on the arm surface. The red line represents the estimated normal direction at individual points on the trajectory.

3.4.4 Confidence-based Probe Orientation Correction

US confidence map was originally introduced by Karamalis *et al.* [61]. They provide a pixel-wise metric to assess the image quality by computing the loss of emitted US signals by the transducer. Regarding the computed confidence maps of B-mode images, a probabilistic map $C \in \mathbb{R}^2 \rightarrow [0, 1]$ is created (see Fig. 3.14). The top white pixels (1) represent the strongest signal and the bottom black pixels (0) represent no US signal arrived. US confidence map has been employed for different aims, e.g., optimizing contact force [53] and tracking target anatomy [62].

---

[6] https://pointclouds.org/

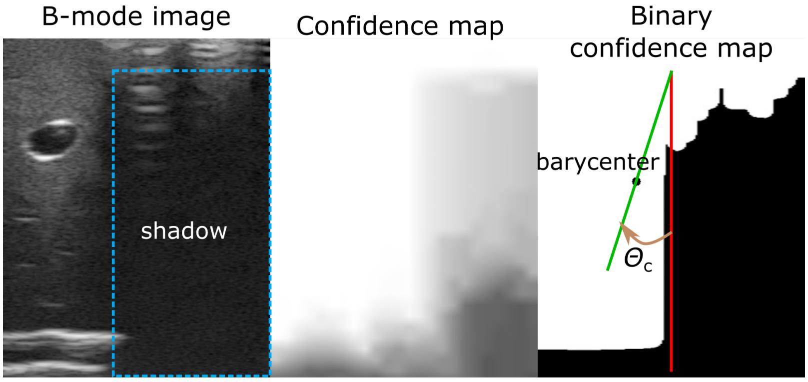


Fig. 3.14. Confidence-based orientation correction. The left, middle and right images are B-mode image, computed confidence image, and binary confidence map. In this case, computed correction angle is $\theta_c = 10.4°$.

In this study, the computed scanning trajectory may not be perfect for scanning due to the error of both hand-eye calibration and surface registration (ICP). The use of a human-like arm phantom with an uneven surface aggravates the negative influence of such errors. More specifically, the US probe may not be able to fully contact the arm surface, which will result in shadow in the imaging (see Fig. 3.14). To properly display the anatomy under the shadow, a confidence-based orientation correction approach is proposed. To this end, the confidence map is further transferred into a binary image by applying a probabilistic threshold (T_{com}). This emphasizes the difference between the well-contacted part and the non-contact part in the resulting images. Then, the weighted barycenter point ζ_c can be calculated as follows:

$$\zeta_c = \frac{1}{\aleph_c} \sum_{0}^{H_{us}} \sum_{0}^{W_{us}} C(h, w) \, [h, w]^T \tag{3.9}$$

where $\aleph_c = \sum_{(h,w)\in\Omega} C(h, w)$ is the accumulated confidence value over the entire image (Ω).

Then, the in-plane rotational adjustment angle θ_c can be calculated by connecting the barycenter and the top center point of image view (see Fig. 3.14). To ensure that the probe is fully in contact with the object surface, multiple in-plane orientation corrections may be required at one position. To reduce the time consumed at each position, the pose (R_p) of the remaining N_{up} points in the trajectory are updated as follows:

$$R_p^{'}(i_c + i) = R_p(i_c + i) R_x(-\eta_i \theta_c) \tag{3.10}$$

where i_c represents the iteration of current position in the trajectory, $i = 1, 2, .., N_{up}$, R_x is the rotation matrix around X direction of probe coordinate system, $\eta_i = \frac{d^2_{N_{up}-i+1}}{\sum_{j=1}^{N_{up}} d_j^2}$ is the weight coefficient used to update the next N_{up} positions.

3.4.5 Movement Identification and Compensation

Movement Monitor System

To effectively detect the potential movement (expected or unexpected) during the scanning, the UNet-VGG16 architecture [104, 107] is employed to segment the arm from RGB images. The U-Net was proposed by Ronneberger [108] for segmentation tasks based on Fully convolutional networks (FCNs). U-Net consists of two parts: the feature extraction part and the up-sampling part with skip connections. Besides, VGG-Net confirmed that a smaller kernel size and a deep CNN can improve model performance [109]. To address the issue of limited data, the VGG-16 architecture was employed as the encoder layer of U-Net structure [107]. With a well-trained VGG-16 model on imageNet data set [110], the training time of the used UNet-VGG16 can be significantly reduced.

To train an UNet-VGG16 model to efficiently segment the arm in this setup, 270 images were collected by changing both the camera position and the environments, e.g., different backgrounds, occlusion caused by the robotic arm and US probe, and with/without humans in the field of view. Then, the arm was manually annotated on all images using LabelMe software [111]. To augment the data for better generalization, the images are randomly flipped and the color are also randomly modified by changing the related parameters, i.e., brightness and contrast. Finally, 3240 images are generated. The ratio between the training and validation data sets is $8:2$. The size of the input images is 512×512. The batch size and the training epochs are 4 and 10, respectively. In addition, the pixel-wise mean squared error L_{mse} is employed as the loss function to train the network.

$$L_{mse} = \frac{1}{N} \sum_{c=1}^{C} \sum_{i=1}^{H_c} \sum_{j=1}^{W_c} [y_c(i,j) - \hat{y}_c(i,j)]^2 \tag{3.11}$$

where $N = C \cdot H_c \cdot W_c$. y and $\hat{y}$ are the annotated image and output of the network.

To validate whether the proposed pipeline is compatible for real scenarios, the second UNet-VGG16 model was trained separately for human volunteers. To avoid articulated movements, only the forearm was selected as target object. The dataset of human arm was generated using the same manner as arm phantom data. Finally, 3560 labeled images from two healthy volunteers (BMI: 25.0 ± 2.8, age: 28.5 ± 1.5) were separated as $8:2$ for training and validation.

After obtaining the binary masks using the trained model, the dice coefficient C_d between the segmented results from the current frame (i-th) and the previous frame ($i-j$-th) is calculated as follows:

$$C_d = \frac{2\left|S_{i-j} \cap S_i\right|}{\left|S_{i-j}\right| + \left|S_i\right|} \tag{3.12}$$

where S_i is the segmented binary results of i-th camera image.

Movement Compensation

If a movement happened during the scanning, the real-time dice coefficient decreases correspondingly. Thus, C_d is used as the metric to identify the object movement. Once C_d becomes smaller than a preset threshold T_{dice}, a signal is emitted to inform the controller to stop scanning via ROS. Based on the experimental results, T_{dice} was empirically set as 0.95 to guarantee both the sensitivity and the robustness of motion detection results in our setup. To properly resume the sweep from the breaking point, the transformation between previous (before the movement) and current objects is calculated using ICP (see Fig. 3.11). Here, the point clouds of the arm surface are created based on the mask of RGB images. After properly assigning the depth information to the pixels located inside the bounding box of the mask, a raw point cloud can be generated. To extract the arm, the plane segmentation algorithm is applied to remove the points under the table surface and an additional depth filter is employed to remove the points higher than a certain value in Z_b direction of frame $\{b\}$. To validate whether the registration results are good enough for continuing the scanning from the break point, an ArUco marker is rigidly placed on the table. To quantitatively access the registration results, the motion compensation error e_{mc} is computed as follows:

$$e_{mc} = \left\| \mathbf{P}^{'}_{ar} - (\mathbf{R}_{mc} \cdot \mathbf{P}_{ar} + \mathbf{V}_{mc}) \right\| \tag{3.13}$$

where $\mathbf{P}_{ar}$ and $\mathbf{P}^{'}_{ar}$ are the positions of the ArUco marker before and after the movement, respectively, while $\mathbf{R}_{mc}$ and $\mathbf{V}_{mc}$ are the computed rotation matrix and translation vector.

The RUSS automatically resumes the scanning from the break point if e_{mc} is small enough ($< 1\ cm$). Otherwise, the RUSS automatically ends the sweep. Compared with the state-of-the-art marker-based approach [6], the use of only camera images makes it more suitable for real clinical routine without the need of carefully configuring the markers on various patients.

Fine Adjustment Algorithm

To well stitch the two or multiple sweeps when movement happens during the scanning, a fine adjustment procedure is further developed to displaying a complete and accurate 3D anatomy. To achieve this, the probe pose of the last frame in the first sweep (before movement) (T^{la}_{be}) and the probe pose of the first frame in the second sweep (after movement) (T^{fi}_{af}) are used. To overlap these two frames, the tracked probe pose ($\mathbf{T}_{be}$) of the first sweep can be updated as follows:

$$\begin{bmatrix} \mathbf{R}^{'}_{be}(i) & \mathbf{P}^{'}_{be}(i) \\ 0 & 1 \end{bmatrix} = \left[(\mathbf{T}^{fi}_{af})^{-1} \mathbf{T}_{mc} \mathbf{T}^{la}_{be} \right]^{-1} \mathbf{T}_{mc} \mathbf{T}_{be}(i) \tag{3.14}$$

where $\mathbf{T}_{mc}$ is the homogeneous expression of $\mathbf{R}_{mc}$ and $\mathbf{V}_{mc}$ in Eq. (3.13).

In addition, a further translational adjustment in the image plane is performed to overlap the two centroids ($\mathbf{P}^{c}_{be}$ and $\mathbf{P}^{c}_{af}$) of the target vessel from the last frame in the first sweep and the first frame in the second sweep, respectively. The centroid is computed based on the binary mask computed using a well-trained U-Net model as [7]. To train the network, 3369 B-mode images were used. The training details are exactly the same as [7]. Finally, the tracked probe pose is thus further updated:

$$\mathbf{P}^{''}_{be}(i) = \mathbf{P}^{'}_{be}(i) +^{b}_{us} \mathbf{R}(\mathbf{P}^{c}_{af} - \mathbf{P}^{c}_{be}) \tag{3.15}$$

3.4.6 Results

Segmentation Results on Phantoms and Human Arm

The network parameters were optimized using Adam [112] on a single GPU (Nvidia GeForce GTX 1080). The learning rate was set to 10^{-3} at the beginning and reduced by a factor of ten when the loss changes were less than 0.0001 for ten subsequent steps. The performance of the well-trained models on unseen images of arm phantom and human forearms are summarized in TABLE 3.1. The average dice coefficient achieved 0.94 and 0.95 on arm phantom and human forearms, respectively. Besides, the segmentation time for each image only take around $5\ ms$.

Tab. 3.1. Performance of Segmentation Algorithm

Dataset	Dice Coefficient	Time (ms)	Samples
Arm phantom	0.94 ± 0.04	5.3 ± 0.6	633
Volunteers forearm	0.95 ± 0.03	5.4 ± 0.6	704

To demonstrate the performance of the segmentation result, four representative results on arm phantom and two different human arm are shown in Fig. 3.15. Regarding the arm phantom, Fig. 3.15 (a) and (b) display segmentation results form RGB images when the arm surface is without/with occlusion caused by US probe, respectively. Regarding the experiments on volunteers, a normal blue tape was warped around human elbow. The tape servers as boundary of forearm and its position is easily changed and determined based on target anatomies by human operators. The segmentation results on two different volunteers are shown in Fig. 3.15 (c) and (d). Even when the occlusion happens, the forearm boundary is still completely extracted. These results demonstrate that well-trained models has the potential to accurately extract the target objects. The dice coefficient of these four representative results are 0.96, 0.95, 0.97 and 0.96, respectively.

ICP-Based Surface Registration Performance

To demonstrate the performance of the ICP-Based surface registration approach in our setup, a representative experimental result has been shown in Fig. 3.16. The mean squared error (MSE) of distance between corresponding points in the two point clouds ($\mathbf{P}_{ct}$ and $\mathbf{P}_c$) quickly decreased at beginning and converged ($0.2\ mm$) after 25 iterations. The initial alignment is shown in Fig. 3.16 (b), which is computed using multiscale feature resistance approach [106]. Then the result after 10 and 40 iterations are depicted in Fig. 3.16 (c) and (d). It can be seen that the point cloud obtain using an RGB-D camera can be well registered to a preoperative template (MSE is $0.16\ mm$).

Considering the potential occlusion in a real scenario, the extracted point cloud from an RGB-D camera $\mathbf{P}_c$ could be incomplete. To further validate the robustness of the ICP-based Registration, $\mathbf{P}_c$ is cropped by a plane being orthogonal to the first principal direction computed using PCA. To investigate the influences caused by various levels of occlusion, $\mathbf{P}_c$ is cropped by the plane at 10%, 20% and 40% on the first principal direction [Fig. 3.17 (a), (b) and (c), respectively] and two planes at 10% and 90% [Fig. 3.17 (d)]. In all these four setups, good

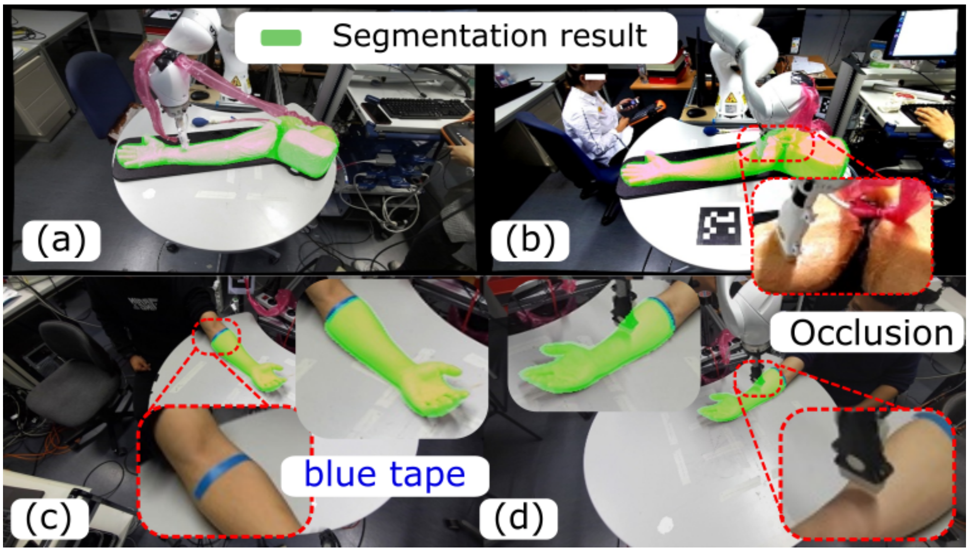


Fig. 3.15. Segmentation performance on unseen images. (a) and (b) are the results on unseen images without and with occlusion, (c) and (d) are the results on unseen images from two different volunteers. The computed dice coefficients for (a), (b), (c) and (d) are 0.96, 0.95, 0.97 and 0.96, respectively.

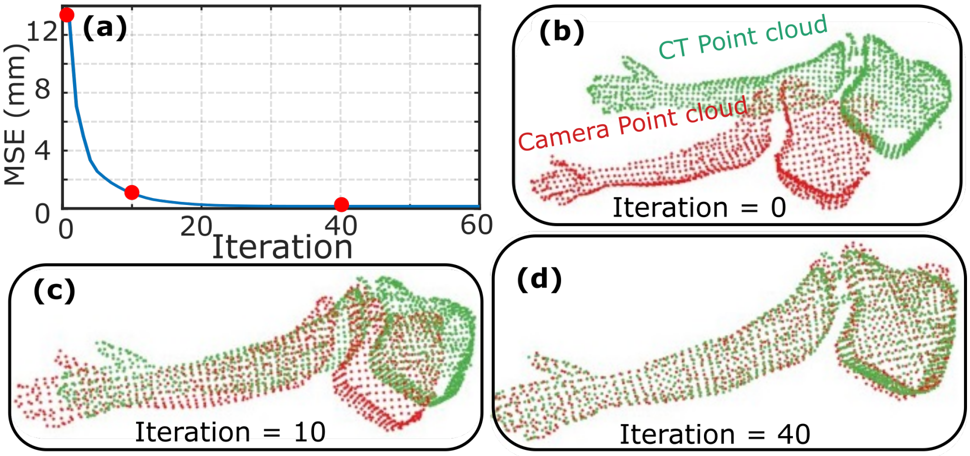


Fig. 3.16. Surface registration results. (a) is the mean squared error (MSE) of distance between corresponding points in the two point clouds, (b), (c) and (d) are the results when the iteration is 0, 10 and 40, respectively. The sizes of preoperative $\mathbf{P}_{ct}$ and camera-based $\mathbf{P}_c$ are 1379 and 925, respectively.

results (Fig. 3.17) are achieved after a certain iterations (64, 68, 83, and 168, respectively). The final MSE are 0.3, 1.6, 4.4 and 2.7 mm, respectively.

Performance of Confidence-based Trajectory Correction

Due to the error of hand-eye calibration and CT-to-object registration, the computed trajectory may result in non-optimal contact during the scan. This reduces the quality of US imaging, like introducing a shadow in the image view as shown in Fig. 3.18 (b). Thus, a confidence-based orientation correction is developed to guarantee good contact condition between the probe and the arm surface during the scan. To demonstrate the performance of the confidence-based correction, two sweeps with an exactly same initial trajectory are performed. The representing results are depicted in Fig. 3.18.

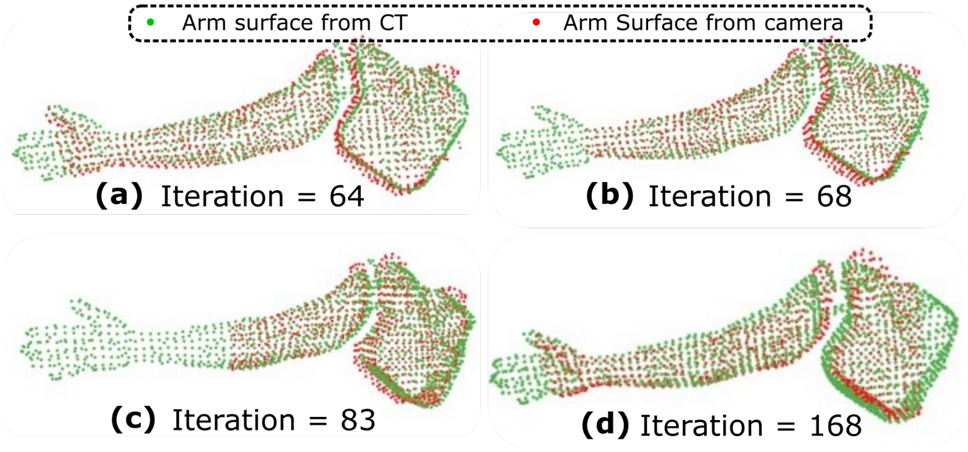


Fig. 3.17. Surface registration results between preoperative point cloud and incomplete camera point cloud. In (a), (b), and (c), the camera point cloud $\mathbf{P}_c$ is cropped from 10%, 20% and 40%, respectively, on the first principal direction. In (d), $\mathbf{P}_c$ is cropped by two planes at 10% and 90% on the first principal direction.

It can be seen from Fig. 3.18, part of the target artery is out of the imaging view for the case without correction. This is represented by the red part in 3D and white circle (incomplete) in 2D [Fig. 3.18 (a) and (b)], respectively. In the case involving correction, once a shadow is detected using the confidence map, an in-plane adjustment angle θ_c is computed and the correction is automatically carried out by the robot. After such correction, the entire artery geometry has been successfully visualized in the imaging view [see Fig. 3.18 (d)]. In addition, since the poses of the remaining positions in the trajectory are also updated using Eq. (3.10), the target vessel is completely visualized inside the imaging view [see Fig. 3.18 (d)]. This result demonstrates that the proposed confidence-based orientation can help to improve the visibility of the target object during the sweep.

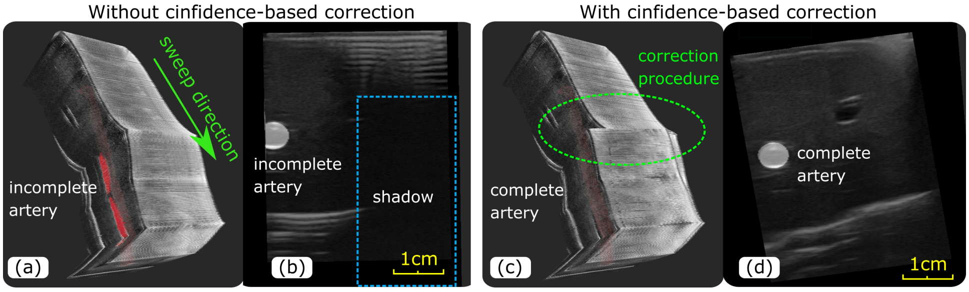


Fig. 3.18. Performance of the confidence-based orientation correction algorithm. (a) and (c) are the 3D results obtained without and with confidence-based correction. (b) is the B-mode images where the target artery geometry is partly out of the view, while (d) is the corrected result where the artery is completely displayed in the imaging view.

Movement Compensation Results

The compensation is done by two successive procedures, 1) surface registration-based compensation and 2) the fine adjustment process using both robotic tracking information and corresponding B-mode images. To qualitatively and quantitatively validate the compensation algorithm, a complete artery of an arm phantom is used as the target object (total length is around $440\ mm$).

1) Performance of Surface Registration-based Compensation: To quantitatively assess the surface registration performance, the phantom arm together with the used table was randomly moved inside of a rectangle ($110mm$) and the maximum rotation angle variation was 80°. To assess the performance of the surface registration-based compensation, an ArUco marker is physically placed on the flat table (see Fig. 3.12). The real-time pose (position and orientation) of the ArUco marker can be computed using a ROS package[7]. Thus, the motion compensation error e_{mc} is computed by Eq. (3.13).

Since the arm surface could be partly occluded by the used robotic manipulator [Fig. 3.15 (b) and (d)] during the scanning, the experiments are separately performed when (case 1) the robotic arm is out of the view and (case 2) is in contact with the phantom arm surface. The experiments were repeated 20 times for each case. The final results are shown in Fig 3.19. The e_{mc} (±SD) of case 1 and case 2 are $6.1 \pm 1.6\ mm$ and $6.9 \pm 1.6\ mm$, respectively. The results obtained when there is an occlusion in the camera view (case 2) is slightly larger than the results of the ideal case (case 1). Besides, based on a t-test (probability $p = 0.14 > 0.05$), it was concluded that no significant difference exists between the two experimental sets. Considering our application, since $6\ mm$ is much smaller than the probe width $37.5\ mm$, the surface registration-based compensation algorithm is promising to be used for compensating a large movement that happens during the scanning. Thereby, a complete geometry of the anatomy can be computed as in Fig. 3.20 (b).

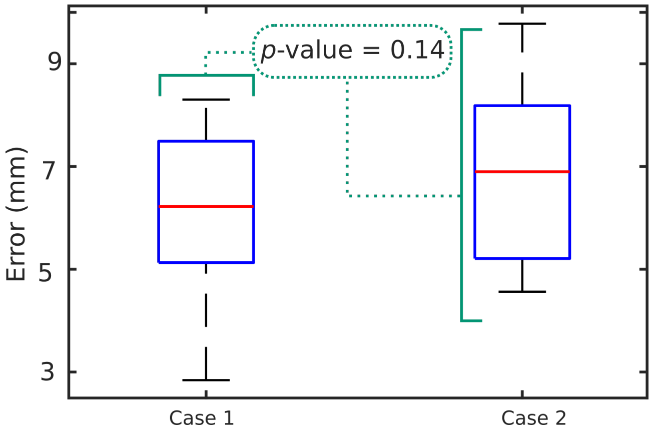


Fig. 3.19. Performance of the surface registration-based compensation algorithm. Case 1 and case 2 represent the situations where the phantom arm surface is not and is occluded by the robotic manipulator, respectively.

2) Performance of Fine Tuning Procedure: After surface registration-based compensation, two partitions have been stitched together. However, due to the error e_{mc}, there is a significant gap and dislocation at the connected part as in Fig. 3.20 (b). To further address this issue, two-step fine adjustment procedures are carried using Eq. (3.14) and (3.15) based on the both robotic tracking information and B-mode images. The tracking information is used to fully overlap the last frame of the first partition and the first frame of the second partition. Then an image-based in-plane adjustment is performed to overlap the two centroids of the target artery in the overlapped two frames. The 3D result after fine adjustment is shown in Fig. 3.20 (c), where the stitching has been successfully compensated. This will further enable autonomous diagnosis.

---

[7] https://github.com/pal-robotics/aruco_ros

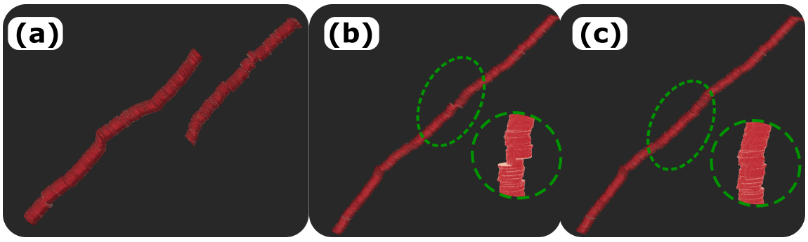


Fig. 3.20. Performance of compensation method. 3D images of a vessel (a) without any compensation, (b) with surface registration-based compensation, and (c) further with a step-wise fine adjustment procedure.

3.4.7 Discussion

Precise repositioning of US probe is one of the crucial techniques for guaranteeing accurate and complete geometry of target anatomies when the object's pose is changed during scanning. This study first proposed a systematic pipeline for RUSS to monitor and further compensate for the potential object motion only on a depth camera. The proposed motion-aware RUSS performs boldly in our setup of experiments (on a human-like arm phantom). However, there are still some limitations that need to be discussed. First, in this work, we have so far considered the rigid motion of object. The articulated motion and deformation of object are existing as well for US scanning. However, the articulated motion could be consider as multiple rigid motion around joints. Regarding the force-induced deformation, one can apply stiffness-based approach to achieve zero-compress volume [5]. In addition, the ICP-based compensation method was designed for visible object motions during US scans. The tremor or vessel motion [103] have not been considered. However, due to the use of the compliant controller, the small and fast physiological motion can be partially adapted by the controller. Thus, compared with free-hand US acquisition, the effect of such small motions is not aggravated in our setup. At this moment, the presented RUSS can be used for the potential applications on rigid body parts, i.e., forearm, upper arm, thigh, shank or whole chest. By further considering articulated motion and integrating deformation correction approach [5] and non-rigid registration [113, 114] in the future, the system will be more robust to be used in regular US examinations.

3.5 Conclusion

In this chapter, we describe the vision-based motion-aware system aiming to achieve accurate and consistent 3D images of target anatomies when an object motion is in presence during US scanning. In this way, the proposed approaches enable sonographers to adjust objects' position and rotation better visualize the underlying objects. This is particularly useful for the anatomies with long structures due the to limited working space of the robotic manipulator. Thus, the proposed framework possesses the advantages of free-hand US (flexibility) and robot US (accuracy and stability) at the same time.

To validate the novel idea of motion-aware RUSS, the **passive marker based approach** is validate for a simplified phantom environment in Sec. 3.3. The results show that the proposed method can accurately compensate a translation movement ($3.1 \pm 1.0\ mm$) and rotation movement ($2.9 \pm 0.7\ mm$) for movements of up to $200\ mm$ and rotations of up to $40°$. In addition, to further explore the effective pipeline in real scenarios, the **markerless approach** is developed with a human-like arm phantom with uneven surface, as Sec. 3.4. To achieve the autonomous manner, the scanning trajectory is extracted from a CT atlas. Compared with the marker-based approach described in Sec. 3.3, a markerless approach is more convenient for operators without the requirement for carefully configuring the markers for individual patients. In addition, to display 3D objects without significant stitching gap as [6], a step-wise fine adjustment procedure is carried out based on both tracking data and B-mode images [see Fig. 3.20 (c)]. Such improvement we believe can make RUSS more robust and thus, bring them closer to clinical use. Furthermore, we consider that it enables the research topic of automatic examination and diagnosis, which is significantly meaningful in pandemic and undeveloped countries and areas.

4

Deformation-Aware Robotic Ultrasound Imaging

4.1 Overview and Publications

This chapter presents the contributions of this thesis regarding deformation-aware robotic US imaging. Due to the inherited characteristic of US imaging, a certain force is required to be applied to achieve optimal acoustic coupling for good visibility of the target anatomies. Due to the exerted pressure, the shape distortion of visualized tissue structures is inevitable, in particular for soft tissues such as superficial blood vessels. The pressure-induced deformation will lead to geometrical errors hindering accurate and operator-independent bio-metric measurements. Such deformation has an even larger effect on 3D US volumes as the correct compounding is limited by the inconsistent location and geometry.

To address this challenge, this chapter presents a patient-specified stiffness-based method to correct the tissue deformations in robotic 3D US acquisitions. To obtain the patient-specified model, robotic palpation is performed at various sampling positions on the tissue. The contact force, US images and the probe poses of the palpation procedure are recorded. The contact force and the probe poses are used to estimate the nonlinear tissue stiffness. The images are fed to an optical flow algorithm to compute the pixel displacement. Then the pixel-wise tissue deformation under different forces is characterized by a coupled quadratic regression. To correct the deformation at unseen positions on the trajectory for building 3D volumes, an interpolation is performed based on the stiffness values computed at the sampling positions. With the stiffness and recorded force, the tissue displacement could be corrected. The method was validated on two blood vessel phantoms with different stiffness. The results demonstrate that the method can effectively correct the force-induced deformation and finally generate 3D tissue geometries.

This chapter is organized as follows: In Sec. 4.2, the clinical and methodological background has been summarized. Sec. 4.3 describes the overview of the involved hardware and the control method. Then, the proposed stiffness-based deformation correction is described in Sec. 4.4. The experimental results on a stiff custom-designed phantom and a commercial soft phantom are presented in Sec. 4.5. Finally, the detailed discussion and the conclusion are displaced in Secs. 4.6 and 4.7, respectively.

Substantial parts of this chapter have already been published in the following article and are quoted verbatim.

[5] **Zhongliang Jiang***, Yue Zhou*, Yuan Bi, Mingchuan Zhou, Thomas Wendler, Nassir Navab. "Deformation-Aware Robotic 3D Ultrasound," *IEEE Robotics and Automation Letters (presented at IROS2021)*, vol. 6, no. 4, pp. 7675-7681, 2021 (*equal contribution).

4.2 Introduction

4.2.1 Clinical Motivation

Ultrasound (US) is a widely utilized diagnostic imaging modality for examinations of internal organs. It is also commonly used to obtain the location and geometric information of disease intraoperatively as US imaging is highly available, non-invasive and radiation-free. However, to obtain optimal acoustic coupling of a US transducer and thus achieve good visibility of target anatomies, a certain pressure is required to be applied to the imaged anatomy. Due to the exerted pressure, the shape distortion of visualized tissue structures is inevitable, particularly for soft tissues such as superficial blood vessels (Fig. 4.1). The shape of the cephalic vein continues to compress when the contact force increases. The vein loses its complete lumen when the force increases to $8\ N$. As a result, the distortion can severely obfuscate the geometrical measurements of subsurface targets, e.g., measuring blood vessel diameter for identifying vascular stenosis.

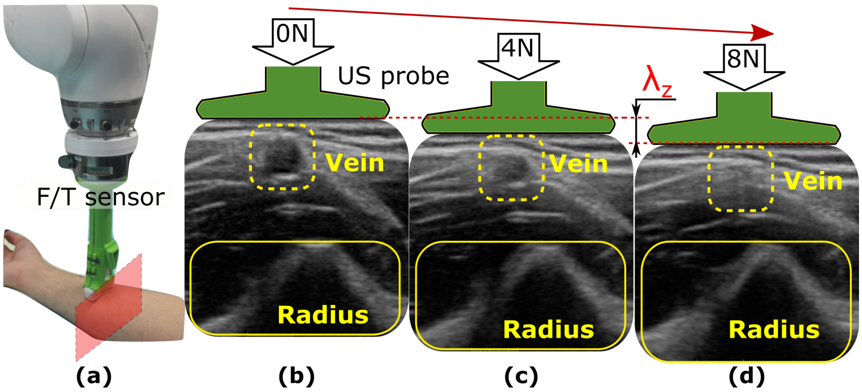


Fig. 4.1. Cephalic vein imaged under three different forces. (a) robotic arm with a force/torque (F/T) sensor, (b), (c) and (d) are the resulting B-mode images acquired when the contact forces are $0\ N$, $4\ N$ and $8\ N$, respectively. The echo of the radius bone is visible below. λ_z represents probe tip displacement in the applied force direction between (b) and (d), here about $6\ mm$. *Reprint from [5] with permission of IEEE.*

4.2.2 Tissue Deformation Correction

For conventional free-hand US examination, it is challenging to maintain a constant force, even for an experienced sonographer. The varied pressure results in different deformations across the individual 2D US slices, which will further hinder the achievement of accurate 3D geometry. Although 3D US imaging can directly display 3D anatomical structures reducing the requirements for sonographers and improving the accuracy of diagnosis [115], the deformation impairs the higher acceptance of such approaches in clinical practice [116].

To address the pressure-induced deformation, Treece *et al.* employed non-rigid image-based registration and position sensing for free-hand 3D musculoskeletal US examinations [117]. This approach mainly focused on axial deformation. Burcher *et al.* built a finite element method (FEM) model to predict the deformation in both axial and lateral directions [118]. The performance of this method relies on the prior knowledge of the tissue properties, which makes it unsuitable for real-world applications. Additionally, Flach *et al.* employed a generic homogeneous representation of tissues to estimate the deformation using FEM [119]. However, tissues are inhomogeneous in reality and their properties vary from one patient to another. To address this drawback, Dahmani *et al.* applied a linear elastic biomechanical model to estimate the personalized mechanical parameters of the tissues along the deformation field [120]. In addition, Sun *et al.* proposed a method to compute a zero-pressure image using an empirical regressive model of US image deformation with respect to the applied force [100]. Nevertheless, the method was developed to correct the deformation in 2D slices. A good regressive model is computed based on a set of forces and corresponding pixel displacements. This method cannot be directly extended to obtain compression-free 3D volumes because it is impractical to acquire paired forces and image deformations in a dense sampling.

Recently, since robotic manipulators are much more accurate and stable than human operators, robotic US systems (RUSS) have become a promising solution toward automatic US scans [7, 11, 42, 56]. To accurately control the probe orientation, Jiang *et al.* developed a method to align the probe centerline with the normal direction of a tested object using force data [9]. Besides, Jiang *et al.* proposed a motion-aware 3D RUSS to combine the flexibility of the free-hand US and the accuracy and stability of robot [6]. In addition, unlike the free-hand manner, the probe pressure-induced indentation to the tissues is accurately controlled to be homogeneous along the entire US volume for RUSS. This is achieved by using hybrid force/position controllers to maintain a certain contact force during the scanning [42]. However, homogeneous deformations still exist in all slices. To further obtain zero-pressure 3D volumes, Virga *et al.* applied a 4th-order polynomial function to regress the force-dependent deformations and further propagate the deformation field at sparse sampling points to the whole scanning direction [116]. Nevertheless, five pixels are required to be manually selected on the first frame. In addition, the method takes around $186\ s$ to compute the deformation field at one location and 15 points need to be selected along a sweep path of $70\ mm$. These drawbacks hinder the acceptance of this method in clinical practice.

4.2.3 Contributions

In this work, a stiffness-based deformation correction method, incorporating image pixel displacements, contact forces and nonlinear tissue stiffness, is proposed to recover a zero-compression 3D tissue geometry from the deformed data recorded during robotic scans. To obtain patient-specific regression models, robotic palpation was performed at sampling positions. Since the tissue compression leads to the increase of tissue stiffness, the nonlinear tissue stiffness was modeled as a 2nd-order polynomial function to the probe displacement in the applied force direction. An optical flow algorithm was employed to compute the two-dimensional pixel displacement in US images acquired under different contact forces. With such displacement, a coupled quadratic regression model with respect to the pixel position, the contact force and the tissue stiffness was computed as an optimization issue for the sampling position. Since the tissue stiffness is the key factor affecting the deformation, the optimized regression model for the initial sampling position can be quickly propagated to other positions on the trajectory by substituting the estimated local stiffness. This speeds up the process to obtain compression-free 3D volumes. The method was validated on two vascular phantoms with significantly different stiffness.

4.3 System Overview

4.3.1 Hardware Overview

The system is comprised of three main components: a robotic manipulator (KUKA LBR iiwa 7 R800, KUKA Roboter GmbH, Germany), a linear US probe (CPLA12875, Cephasonics, USA) and a Gamma force/torque (F/T) sensor (ATI Industrial Automation, USA). The F/T sensor is attached to the robotic flange. The US probe is mounted on the other side of the F/T sensor using a custom-designed holder. The robotic system is controlled using a Robot Operating System (ROS) interface [48]. The control commands are exchanged at $100\ Hz$. The whole system is depicted in Fig. 4.2.

The contact force between the probe and the object is recorded using a data acquisition device (FTD-DAQ-USB6361, National Instruments, USA) and further published to ROS with accurate timestamps. Besides, the US images ($50\ fps$) are accessed via a USB interface and visualized using a software platform (ImFusion GmbH, Munich, Germany). The detailed acquisition parameters are set as follows: image depth: $40\ mm$, frequency: $7.6MHz$, brightness: $66\ dB$. To synchronize the US images, forces, and probe poses, the US images are also published to ROS with timestamps.

To validate the performance of the proposed method to recover zero-compression images from deformed images on different tissues, two blood vessel phantoms with different stiffness are employed (Fig. 4.2). The one with higher stiffness is custom-made by gelatin powder ($175\ g/L$) and paper pulp (3-5 g/L). The paper pulp is used to mimic the unstructured artifacts of human tissues in US images. In addition, to increase the durability of the gel phantom (over one month), a liquid disinfectant is mixed with water ($1 : 9$). The second phantom (blue phantom, CAE, FL, USA) is softer than the custom-designed one. The phantoms with significantly different stiffness are used to validate whether the proposed method is able to quickly extrapolate its optimized correction model to different tissues without any prior knowledge.

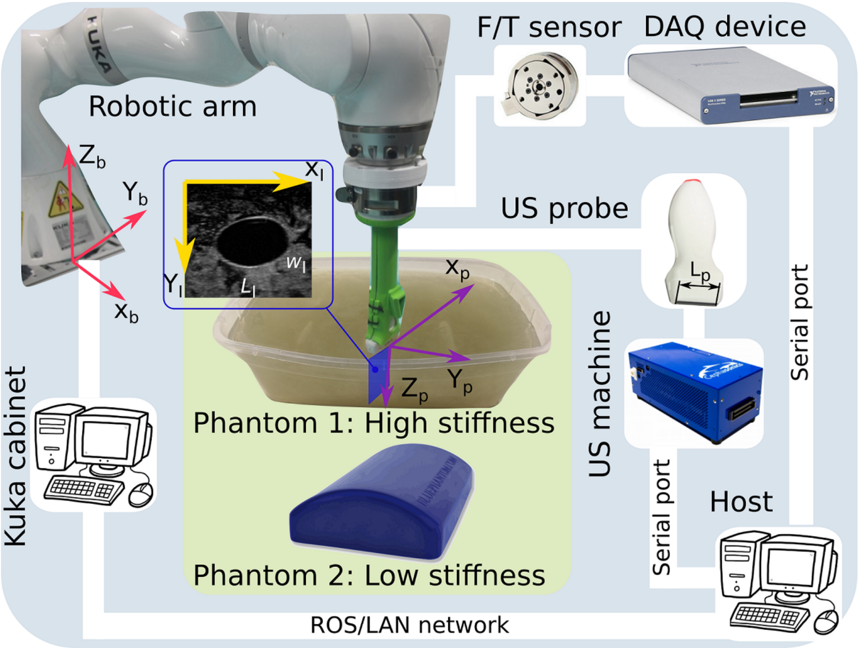


Fig. 4.2. System overview. *Reprint from [5] with permission of IEEE.*

4.3.2 System Calibration

To generate a 3D US volume from the tracked 2D B-scans, the spatial calibration should be employed to project the pixels to voxels in 3D space. There are three coordinate frames involved: robotic base frame $\{b\}$, US probe tip frame $\{p\}$, and B-mode imaging frame $\{I\}$ (see Fig. 4.2). Since Z direction of frame $\{I\}$ is set in the perpendicular direction of the 2D image plane, a homogeneous representation of pixel position in frame $\{I\}$ can be written as ${}^{I}\mathbf{P} = [x, y, 0, 1]^{T}$. Accordingly, 2D pixel positions ${}^{I}\mathbf{P}$ are projected into 3D space using Eq. (4.1).

$$ {}^{b}\mathbf{P} = {}^{b}_{p}\mathbf{T}\, {}^{p}_{I}\mathbf{T}\, {}^{I}\mathbf{P} \tag{4.1} $$

where ${}^{j}_{i}\mathbf{T}$ is the transformation matrix used to transfer the position from frame $\{i\}$ to frame $\{j\}$.

The transformation matrix ${}^{b}_{p}\mathbf{T}$ is computed based on the kinematic model and the 3D geometry parameters of the 3D printed probe holder. In addition, considering that the origin of frame $\{p\}$ is set to the central point of the probe flat tip surface, and that the origin of frame $\{I\}$ is on the top left side, the matrix ${}^{p}_{I}\mathbf{T}$ is represented as follows:

$$
{}^{p}_{I}\mathbf{T} = \begin{bmatrix} \frac{L_p}{L_I} & 0 & 0 & -\frac{L_p}{2} \\ 0 & 0 & -1 & 0 \\ 0 & \frac{D_I}{W_I} & 0 & \varepsilon \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4.2}
$$

where L_I and W_I are the length and width (in pixels) of the 2D images (see Fig. 4.2). D_I is the physical depth of the US scan, namely $40\ mm$. L_p is the physical length of the deployed US elements, here $L_p = 37.5\ mm$. ε is used to represent the small distance from probe frame origin to image frame origin determined by the US elements configuration.

4.3.3 Control Architecture

To ensure the image quality and human safety, a compliant controller is essential for robotic US acquisition [48]. The used Cartesian compliant control law is defined in Eq. (4.3)

$$
\tau = J^T[K_m(x_d - x_c) + F_d] + D(d_m) + f_{dyn}(q, \dot{q}, \ddot{q})] \tag{4.3}
$$

where τ is the target torque applied to the joint drivers, J^T is the transposed Jacobian matrix, x_d and F_d are the desired position and the target force, x_c is current position, K_m represents the Cartesian stiffness, $D(d_m)$ is the damping term, and $f_{dyn}(q, \dot{q}, \ddot{q})$ is the dynamic model of the robotic arm.

To make the force stable in the direction of the probe centerline, the robot is controlled via a 1-DOF compliant controller and a 5-DOF position controller. This behavior is achieved by assigning high stiffness values to the DOF that will be controlled under position mode. Regarding the compliant direction, the stiffness is set to a value in $[125, 500]\ N/m$ for different human tissues [48]

4.4 Pressure-Induced Deformation Correction

The overall pipeline proposed to compute compression-free 3D volumes based on multiple sensor information and a robotic platform has been shown in Fig. 4.3. To overcome the limitation of the state-of-the-art method [116], the proposed approach is developed based on the tissue stiffness. The use of stiffness allows to quickly propagate an estimated 2D deformation regression from one position to another position, or even extrapolate the optimized regression model to other tissues with totally different stiffness profiles.

Being beneficial from using a robotic platform, palpation can be automatically performed, and paired forces, probe poses and US images are recorded. Afterward, the local stiffness with respect to the contact force and tissue position is estimated. The unknown parameters of the regression model are optimized based on the pixel movements, extracted using optical flow technique, in the axial and lateral direction, respectively. Considering the tissue properties are not homogeneous, robotic palpation is performed multiple times at different positions. Instead

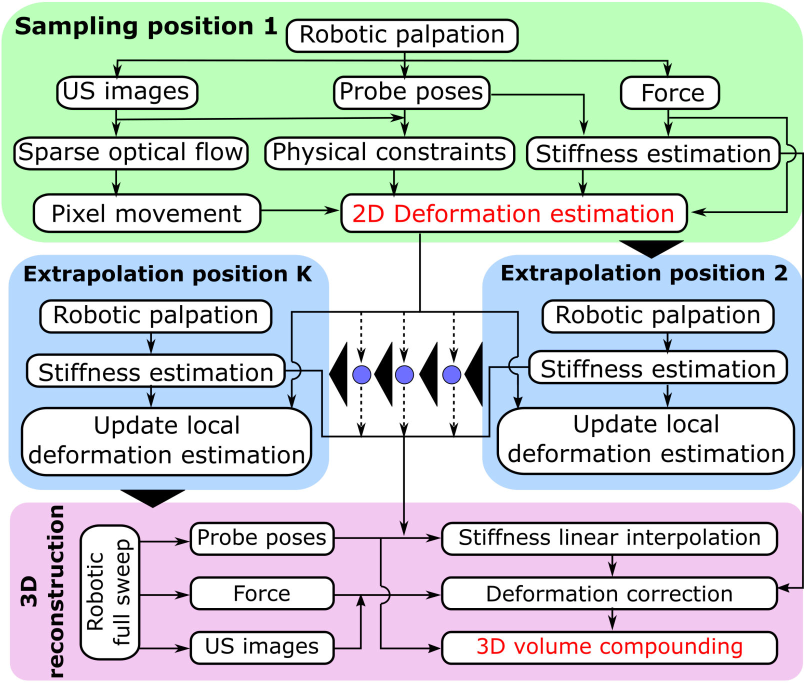


Fig. 4.3. Pipeline to correct the force-induced deformation and reconstruct 3D US volumes. At least two palpation procedures should be performed at different positions to obtain the zero-compression volume. *Reprint from [5] with permission of IEEE.*

of running multiple optimization procedures, the proposed method can update the deformation regression by substituting the local stiffness profiles in the optimized model. Finally, to obtain 3D volumes, a robotic sweep is performed along the planned path. The deformations of the US images acquired at unsampled positions are corrected by the approximated local stiffness computed based on the sampled positions with linear interpolation. Based on the corrected 2D images and probe poses, the 3D volume can be compounded. The detailed steps are described in the following subsections.

4.4.1 2D Deformation Estimation

To accurately visualize the tissue geometry and enable better diagnosis, a 2D deformation field is computed using two regression functions with respect to tissue stiffness and coupling pixel displacements. The tissues with high stiffness can resist deformations, while the tissues with low stiffness will suffer larger deformations during scans. However, most model-based deformation correction methods directly compute the pixel displacements with respect to the contact force (F_c) rather than to the stiffness (k) [100, 116, 117]. Since F_c is applied externally, it cannot reflect tissue's deformation-resistant capacity. Therefore, the computed deformation field cannot be propagated on other positions or other tissues, which limits the force-based approach for building zero-compression 3D volumes. To propagate deformation fields for 3D volume, an additional 3D inpainting technique is required as in [116]. Yet, this technique takes too much time to compute the 2D deformation field. It reported that 15 points are needed to build a good 3D volume of the sweep lengthen $70\ mm$, and the computation for each point takes $186\ s$ on average. Thus, here a stiffness-based displacement model is built. Since stiffness reflects tissue properties, the proposed approach allows to quickly adapt optimized regressions from sampled positions to unseen positions and even to other tissues.

Dynamic Tissue Stiffness Estimation

Considering a real scenario, tissue stiffness also varies depending on the compression situation. The stiffness becomes large when large deformations occurred. As an example, the recorded force and probe tip displacement λ_z (Fig. 4.1) on the custom-designed stiff and the commercial soft phantom are shown in Fig. 4.4. During the robotic palpation procedure, the contact force is slowly increased $(0, 30\ N)$ for the stiff phantom and $(0, 16\ N)$ for the soft one. It can be seen from Fig. 4.4 that the samples acquired on the stiff phantom are distributed around a linear function while the samples acquired on the soft one are distributed around a nonlinear function. To qualitatively analyze the performance of the stiffness regression models, R-Squared (R^2) is measured. A quadratic function is already able to well explain the samples obtained on both stiff and soft tissues $(R^2 = 0.998)$ while R^2 is only improved by 0.001 using a cubic function. Therefore, a quadratic function is employed to capture the flexibility of elastic tissues as follows:

$$F_c = c_1\lambda_z^2 + c_2\lambda_z + c_3 \tag{4.4}$$

where c_1, c_2 and c_3 are the constant coefficients of the quadratic stiffness regression model.

The dynamic stiffness k_d according to current deformation is computed as the derivative of Eq. (4.4).

$$k_d = 2c_1\lambda_z + c_2 \tag{4.5}$$

Pixel-wise Displacement Regression

To characterize tissue deformation, pixel tracking is performed between US images acquired under different pressures using optical flow. Considering the accuracy and the cost of time, Lucas-Kanade approach [121] is employed to estimate the pixel displacements in both lateral and axial directions $\mathbf{D}_p^{op} = \left[D_p^x, D_p^y\right]^T$. Considering the pixel displacement d_p is a coupled result over different factors, a high coupled polynomial regression over pixel position (x, y), contact force F_c and the estimated tissue stiffness k_d are built. Since the deformation can be calculated using F_c and k_d, $h = \frac{F_c}{k_d}$ is defined to estimate d_p as follows:

$$\mathbf{d}_p(x, y, \frac{F_c}{k_d}) = \begin{bmatrix} \mathbf{K}^x \\ \mathbf{K}^y \end{bmatrix} \mathbf{M}_v(x, y, \frac{F_c}{k_d}) \tag{4.6}$$

where $\mathbf{K}^x \in R^{1\times 10}$ and $\mathbf{K}^y \in R^{1\times 10}$ are unknown parameters of the pixel regressions in lateral and axial direction, respectively. $\mathbf{M}_v(x, y, h) = [x^2,\ y^2,\ h^2,\ xy,\ xh,\ yh,\ x,\ y,\ h,\ 1]^T$ includes all variable combinations (2nd and 1st order) representing the coupling effect as [100].

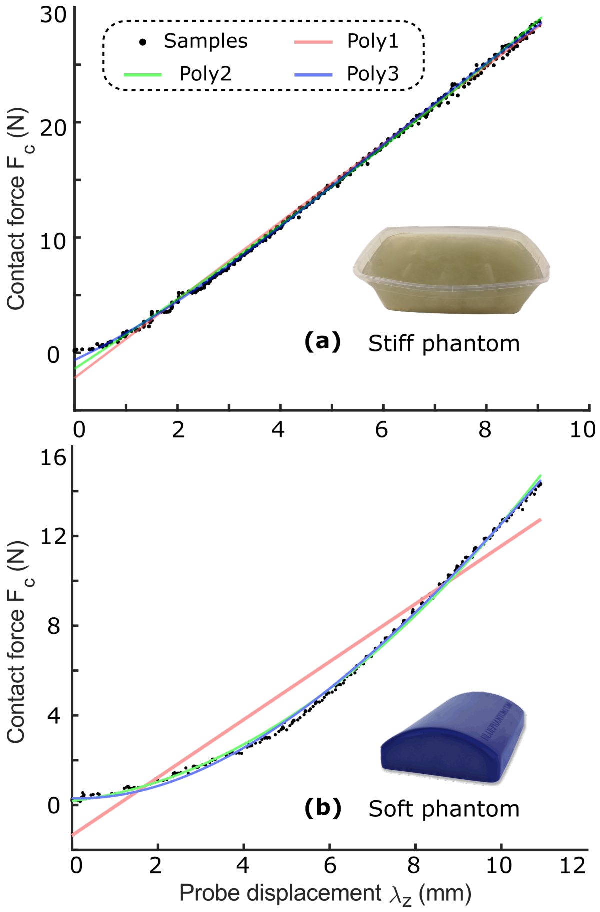


Fig. 4.4. Regression model for tissue stiffness. Poly1, Poly2 and Poly3 represent first, second and third order polynomial. The computed stiffness (Mean±SD) for the two phantoms are 3237 ± 56 N/m and 1489 ± 617 N/m, respectively. *Reprint from [5] with permission of IEEE.*

However, the stiffness of soft tissues varies a lot with respect to the applied force as Fig. 4.4 (b). Eq. (4.6) is only able to depict the deformation of stiff tissues. To make the proposed method also work on soft tissues, the idea of calculus is employed to formalize the final regression model for each pixel as follows:

$$\mathbf{D}_p(F_c) = \int_0^{F_c} \mathbf{d}_p(x, y, \frac{F}{k_d})dF \tag{4.7}$$

where $\mathbf{D}_p(F_c)$ is the cumulative result of $\mathbf{d}_p$ computed for small force intervals, k_d is seen as a constant in each interval.

Besides the dynamic stiffness, this work also employs the probe displacement λ_z (see Fig. 4.1) as boundary constraints to better recover the geometry of the scanned tissue. With a larger λ_z, US imaging will visualize deeper tissues. This means the bottom pixels of the images acquired using small force will be moved toward the top when F_c is increased. Due to the contact between the probe and object, the movement of the superficial layer is mainly in the axial direction. In addition, based on the optical flow results, the displacement of the pixels located in the low part also mainly happens in the axial direction, especially when a large force is applied. Thus, the physical constraints can be written as follows:

$$\begin{aligned} \mathbf{D}_p^b(x, 0, F_c) &= [0, 0]^T \\ \mathbf{D}_p^b(x, L_I, F_c) &= [0, \frac{\lambda_z}{L_T} L_I]^T \end{aligned} \qquad 1 \le x \le W_I \tag{4.8}$$

where L_I and W_I represent the pixel size (length and width) of the US images, and L_T is the thickness of the flexible layer.

Then, the unknown parameters $\mathbf{K}^x \in R^{1\times 10}$ and $\mathbf{K}^y \in R^{1\times 10}$ can be optimized using Eq. (4.9)

$$\min_{\mathbf{K}^x,\ \mathbf{K}^y} \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} ||\mathbf{D}_p(x^j, y^j, \frac{F_c^i}{k_d^i}) - \mathbf{D}_p^m(x^j, y^j, F_c^i)||^2 \tag{4.9}$$

where $\mathbf{D}_p^m(x, y, F_c)$ is the measured pixel displacement. It consists of $\mathbf{D}_p^b$ and $\mathbf{D}_p^{op}$. M is the number of paired force values and US images recorded during robotic palpation, N is the number of the characterized pixel positions involved in optical flow and the boundary constraints. To make the regression can be extrapolated to other positions with different stiffness, F_c and the pixel displacement d_p^x and d_p^y have been normalized to $[0, 1]$. Besides, Eq. (4.9) is optimized using ADAM optimizer [112] with a updating step of 0.01.

4.4.2 3D Compression-Free Reconstruction

The pixel displacement regression model at an initial sampling position on the scanned tissue can be computed as described in Section 4.4.1. The US deformations at this position are computed by substituting the measured F_c and λ_z in the regression. With the computed pixel-wise displacements under different forces, the deformation can be compensated. To obtain the zero-compression volume, the correction procedures are carried out for the whole trajectory. To characterize the local feature, N_k equidistant sampling positions are selected on the trajectory. Subsequently, the robotic palpation is performed at all sampling points and the corresponding stiffness can be estimated. Then, the deformation regression for a different position can be updated by substituting new stiffness.

$$\mathbf{D}_p(F_c) = \int_0^{F_c} \mathbf{d}_p(x, y, \frac{F}{k_d^i}) dF \tag{4.10}$$

where k_d^i is the nonlinear tissue stiffness with respect to the probe displacement λ_z at i-th sampling position.

Since dense sampling is impractical for real scenarios, a robotic sweep over the anatomy is performed. Based on the probe poses, the local stiffness for unsampled positions on the trajectory is computed as follows:

$$k_d^{'} = \sum_{i=0}^{N_k} \omega_k^i k_d^i \tag{4.11}$$

where ω_k^i is weight for the stiffness at N_k sampled positions, $\sum_{i=0}^{N_k}\omega_k^i = 1$. ω_k^i is determined using the distance L_{d_i} from an unsampled position to all sampled positions as Eq. (4.12).

$$\omega_k = \begin{cases} \left[\frac{\sum_{i=1}^{N_k} L_{d_i}}{L_{d_1}}, \dots, \frac{\sum_{i=1}^{N_k} L_{d_i}}{L_{d_{N_k}}}\right]_{Normal} & \forall L_{d_j} \neq 0 \\ [0, \dots, \underbrace{1}_{j-th}, \dots, 0] & \exists L_{d_j} = 0 \end{cases} \tag{4.12}$$

Based on the measured force and the estimated stiffness in Eq. (4.11), the deformations at unsampled positions can also be compensated using the optimized pixel regression for the initial position. The estimated compression-free volume for the target tissues is compounded using the paired corrected US images and the corresponding probe poses as [6].

4.5 Results

4.5.1 2D Compensation Results

Validation of the Deformation Correction Method on the Sampling Position

To validate the performance of the proposed method for generating zero-compression US images from deformed ones, robotic palpation was performed three times at the same position on the stiff phantom. The two sets of data are used to optimize the unknown regression parameters $\mathbf{K}^x$ and $\mathbf{K}^y$ in Eq. (4.6). The results are shown in TABLE 4.1 (taking one arbitrary sampling point as an example).

Tab. 4.1. Optimized Regression Parameters

Parameters	Value	Loss
$\mathbf{K}^x$	$[-0.27, -0.01, 0.15, -0.25, -0.24,$ $-0.27, 0.29, 0.03, 0.16, -2.01]e^{-1}$	$2.5e^{-5}$
$\mathbf{K}^y$	$[-1.43, -1.89, -4.62, -0.87, -0.81,$ $3.69, 0.80, -1.91, 2.29, -2.23]e^{-1}$	

The paired images, contact forces and probe displacements recorded from the third palpation at the same position are then used for validation. To demonstrate the correction performance, the results achieved when $F_c = 25\ N$ is shown in Fig. 4.5. The mimic artery geometry has been significantly compressed in the left view (Fig. 4.5). This affects measurement accuracy of the object's geometry. Nonetheless, based on the results shown in Fig. 4.5, the deformed geometry (red line) can be effectively recovered to ground truth (green line). To further quantitatively analyze the result, the dice coefficient is computed and it has been improved from 0.69 to 0.92.

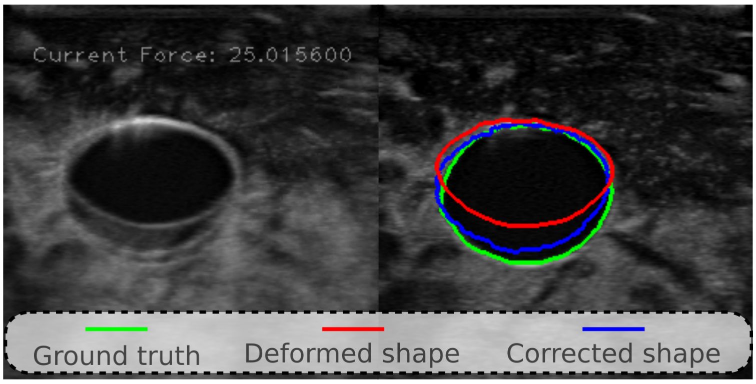


Fig. 4.5. Performance of the proposed deformation correction method for the 2D image obtained under $25\ N$ on the stiff vascular phantom. The deformed image, corrected image and the ground truth acquired when the contact force is zero are overlapped in the left plot. The right image shows the extracted blood vessel boundaries. *Reprint from [5] with permission of IEEE.*

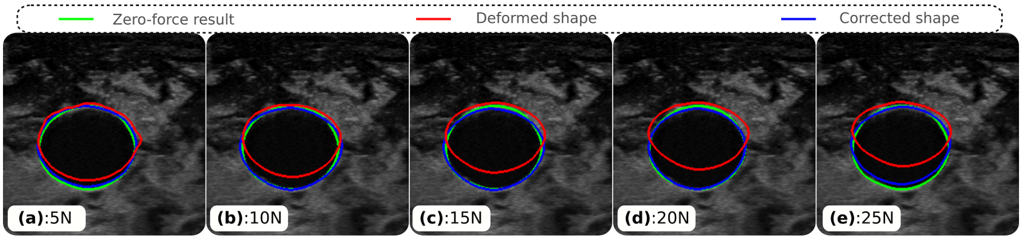


Fig. 4.6. The deformation correction performance on the same tissue at another position using the updated regression model based on local estimated stiffness. The computed dice coefficients for the deformed images and the corrected images are $[0.89, 0.87, 0.81, 0.78, 0.74]$ and $[0.97, 0.97, 0.97, 0.98, 0.94]$, respectively. *Reprint from [5] with permission of IEEE.*

Validation of Propagated Regression Model Computed based on Local Stiffness

Considering the tissue properties are not the same at different positions, multiple palpation procedures are performed at two different positions (larger than $20\ mm$ from the initial position) on the same gel phantom. Some results are shown in Fig. 4.6, which shows that the deformed tissue can be corrected using the updated regression model. The average dice coefficient for the corrected images is 0.96. Besides, more results can be found in this video[1].

Validation of Propagated Regression Model on Other Tissues

To further validate the adaptive ability of the proposed method, a commercial vessel phantom is employed. The average stiffness of the soft phantom ($1489 \pm 617\ N/m$) is only about half of the stiff one ($3237 \pm 56\ N/m$). Besides, the stiffness of the soft phantom also varies much more than the stiff phantom when different contact forces are applied. This makes it challenging to adapt the optimized regression model based on the data acquired on the stiff phantom to an“unseen" soft phantom with such significantly different properties. The results are shown in Fig. 4.7. The deformed vessel geometry (red ellipse) can be compensated as well using the proposed approach, and the corrected result (blue ellipse) is very close to the zero-compression data (green ellipse). The dice coefficients improved from 0.66 to 0.89 (i.e. by 35%). In addition, to demonstrate the advantages of the proposed stiffness-based approach over the existing force-based approach [100], the corrected result (yellow ellipse) obtained

---

[1] https://www.youtube.com/watch?v=MlZtugQ2cvQ

using the force-based approach is also shown in Fig. 4.7. This result is obtained by substituting the force on “unseen" tissues to an optimized quadratic polynomial regression model obtained for a sampled position on the stiff phantom. The corrected result of the propagated force-based model (dice: 0.74) is significantly worse than the result of the proposed approach (dice: 0.89).

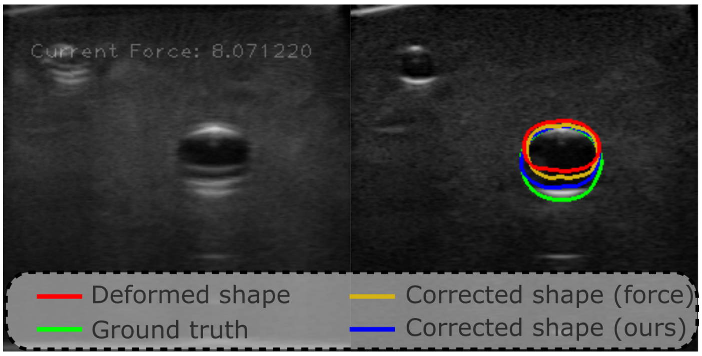


Fig. 4.7. Extrapolation performance on an “unseen" soft phantom. Left: The overlapped result of the deformed image, ground truth and corrected images using the force-based approach [100] and the proposed stiffness-based approach. Right: Vessel boundary detected on the four images. $F_c = 8\ N$.

Finally, to systematically demonstrate the performance of the proposed method on different sampling positions both on stiff and soft tissues, the average dice coefficients for the deformed and corrected results have been summarized in TABLE 4.2. For the stiff phantom, four different sampling positions are employed to validate the performance of the adapted regression models at other positions. At each position, F_c was gradually increased from zero to $25\ N$. To demonstrate the correction performance under various force, the results obtained when $F_c = 5, 10, 15, 20, 25\ N$ are summarized in TABLE 4.2. Then, we further validate the method on the soft phantom. Since the phantom itself is softer, severe deformation happens when $F_c = 8\ N$ (Fig. 4.7). Thus, the typical results acquired when $F_c = 2, 4, 6, 8\ N$ over three sampling positions are summarized in TABLE 4.2. TABLE 4.2 shows that all dice coefficients of the corrected results are over 0.91 while the worst deformed result yields only 0.69.

In addition, the adapted results of force-based approach [100] on the same points were computed as well (TABLE 4.2). Regarding the stiffness phantom, although the results of the proposed approach outperformed the results of the force-based approach in general, both of them achieved good performance. This is mainly because the stiff phantom is homogeneous, which resulting in similar stiffness for different positions on the same phantom. Since the stiffness of the soft phantom is really different from the stiff phantom, the adapted performance of the force-based approach becomes significantly worse than the results obtained using the proposed approach. The ability to quickly and robustly propagate an optimized regression to “unseen" position or even tissues is important to generate a patient-specified correction model in real clinical practice.

4.5.2 3D Reconstruction Results

To obtain 3D volumes of objects, a sweep covering the area of interest is carried out. Considering the variation of tissue properties, N_k sampling positions are selected. The dynamic

Tab. 4.2. Dice Coefficient on Different Tissues and Positions

Results	Stiff phantom					Soft phantom			
	5N	10N	15N	20N	25N	2N	4N	6N	8N
Deformed	0.92	0.86	0.82	0.77	0.72	0.92	0.84	0.75	0.69
Sun *et al.* [96]	0.92	0.96	0.94	0.91	0.90	**0.92**	0.90	0.86	0.82
Ours	0.92	0.96	**0.96**	**0.95**	**0.92**	0.91	**0.92**	**0.92**	**0.91**

stiffness at the sampling position is estimated based on the recorded λ_z. Here $N_k = 4$ is selected on the scan trajectory lengthened $40\ mm$ for the stiff phantom while $N_k = 3$ is used for the soft one, where the path is $60\ mm$. To demonstrate the performance of the proposed stiffness-based regression model on the whole sweep, including many unsampled positions, the 3D volumes of the gel phantom (ground truth, deformed and corrected images) are shown in Fig. 4.8. The ground truth is obtained when the contact force is zero. To achieve this, the phantom is submerged in warm water to avoid air between the probe tip and object surface.

To better visualize and compare the resulting volumes, the views on the axial plane (dash red line) and coronal plane (dash blue line) are displayed in Fig. 4.8. The diameters of the three volumes measured in axial view are similar. However, in coronal view, the length of the orthogonal artery axis (yellow line) measured in the deformed results is less than 20% of the length measured in the ground truth. After applying the proposed deformation correction method, the corrected results have successfully recovered the deformed coronal axis length from being almost identical to the length in the ground truth.

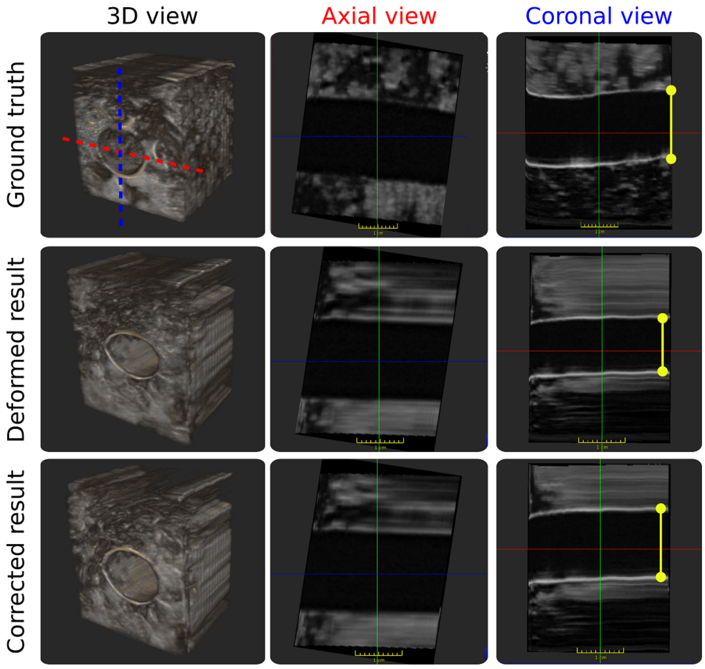


Fig. 4.8. 3D compounding results. The ground truth is recorded when the contact force is zero (phantom is submerged in water). The deformed result is obtained when the force is $15\ N$ on the stiff phantom. *Reprint from [5] with permission of IEEE.*

To validate the accuracy and repeatability of the proposed robotic 3D US system, multiple robotic sweeps are carried out on the same trajectory. To generate the sweep with different levels of deformation, the US sweep is performed with various contact forces. Considering the phantoms' stiffness, five F_c $(5, 10, 15, 20$ and $25\ N)$ are used for the stiff gel phantom and four F_c $(2, 4, 6, 8\ N)$ are used for the soft phantom. To statistically compare the correction performance in terms of force and stiffness, the vessel centroid variance and the cross-section area are introduced as the assessment metrics (see Fig. 4.9). The cross-section area is important because it is the main factor affecting the blood flow velocity and the flow volume pass in unit time. Besides, the centroid is computed for each B-mode image using OpenCV. Then, the centroid variance is computed as the Euclidean distances between the ground truth centroid and the centroids of the deformed and the corrected images. The variance only becomes small when both the long axis and short axis length of the vessel are close to the corresponding values in the ground truth.

Ten images are randomly selected from sweeps acquired under different F_c. The average results of the vessel centroid variance and the cross-section area are shown in Fig. 4.9. It is noteworthy that the centroid variance is significantly reduced after correction in all investigated cases (different forces and phantoms), particularly for soft tissues. The variances for the corrected results are much smaller and more stable than the deformed cases. The maximum variances are $0.78\ mm$ and $2.70\ mm$ for the corrected and the deformed results on the soft phantom when $F_c = 8\ N$. Regarding the stiff phantom, the maximum variances are $1.44\ mm$ and $3.6\ mm$ when $F_c = 25\ N$. These values are larger than those peers obtained on the soft phantom. This is because the absolute geometry size of the artery in the stiff phantom is much larger than that in the soft phantom $(224\ mm^2$ versus $57\ mm^2)$.

Regarding the vessel area, the corrected results are also closer to the ground truth than the deformed results. In the best-case scenario, F_c is small for both phantoms $(2\ N$ for soft phantom and $5\ N$ for stiff phantom) because the pressure-induced deformation is also small when a small force is employed. However, in terms of the difference between the area of the ground truth and the corrected/deformed results, it is significantly reduced from $[17, 34, 47, 59, 66]\ mm^2$ to $[1, 15, 16, 27, 28]\ mm^2$ for the stiff phantom and from $[10, 15, 18, 22]\ mm^2$ to $[-3, 3, 11, 13]\ mm^2$ for the soft phantom. The improvement is over 55% for all cases. Moreover, it can be seen from Fig. 4.9 that, the areas of the corrected results of the images acquired under different forces are very stable. The standard deviation between the areas of corrected images is only $8.8\ mm^2$ and $5.7\ mm^2$ for the stiff and the soft phantoms, respectively. This demonstrates that the proposed system could generate zero-compression 3D US using the proposed deformation correction.

4.6 Discussion

The pressure-induced deformation is a common issue for US imaging of soft tissues. With the proposed approach, accurate and reproducible 3D images, independent of the experience of sonographers, can be achieved by correcting the US deformation. To qualitatively and quantitatively validate the proposed approach, blood vessel is investigated as the target anatomy. However, this study can benefit other applications requiring accurate geometrical measurements as well, such as examining and monitoring breast tumors [122] and soft tissue sarcomas [116]. Besides accurate 3D volumes, this study will also benefit the multimodal image fusion. The two typical clinical applications are image-guided intervention for soft

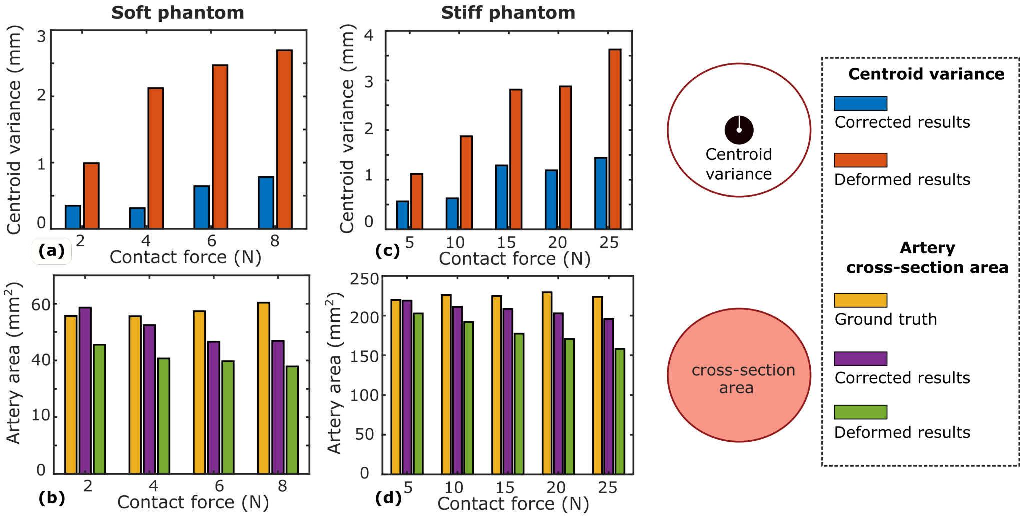


Fig. 4.9. (a) and (c) are the vessel centered variance between the ground truth and corrected and deformed results for the soft and the stiff phantoms, respectively. (b) and (d) are the vessel cross-section area for soft and stiff phantoms, respectively. *Reprint from [5] with permission of IEEE.*

tissues like breasts [123, 124], and imaging-guided orthopedic surgery [9]. For the former one, CT or MR is often used to provide high-resolution anatomies, while US images provide a live view during the intervention. Regarding the latter one, since patients may be moved after acquiring CT or MR images, it is necessary to do the registration between pre-operative images and patients by registering the live US images to the preoperative images. The corrected images could result in better results because it is easier to align the biological landmarks between the zero-compression B-mode images and preoperative images.

The proposed deformation correction approach performs boldly in our setup of experiments. However, there are still some limitations that need to be discussed. First, the proposed method is only able to correct the deformations caused by the force applied on the object surface. This approach is not suitable for correcting deformations of deep organs, such as the liver. These deformations are caused by the coupling of applied external force and patient physiological movement like breathing. In addition, to further improve the accuracy, the probe orientation should also be taken into consideration. This is not implemented here. Alternatively, one can apply methods that enable the probe to be positioned perpendicular to the surface [11]. Finally, in this work, we have so far considered the main in-plane deformation caused by the probe pressure on the surface. Such forces also result in residual out-of-plane deformations. However, since the main component of such deformation remains in-plane and the robotic ultrasound moves orthogonal to the B-mode acquiring all subsequent image planes, we consider this to be a valid approximation of the real tissue deformation.

4.7 Conclusion

In this work, we proposed a stiffness-based deformation correction method in order to achieve zero-compression 3D US images using robotic techniques. This method takes the nonlinear property of tissue stiffness as a key factor in correcting deformation. We obtained this patient-specified property by performing robotic palpation. Our approach enables the rapid adaptation of the optimized regression model to unseen positions by updating the local stiffness. Promising correction results have been achieved on both stiff and soft phantoms at arbitrary sampling positions. Additionally, the experimental results for 3D US acquired under different contact forces demonstrate that the proposed method is also able to recover zero-compression volumes from deformed images. This approach could even enable further techniques such as multi-modal image fusion. The potential clinical applications include examination of soft tissue sarcomas [116], the thyroid gland [125] and blood vessels on limbs.

5

Autonomous Robotic Ultrasound Screening of Tubular Structures

5.1 Overview and Publications

This chapter presents the contributions of this thesis concerning automatic robotic US screening of tubular structures. Due to the high inter and intra-operator variation and a lack of repeatability of US image acquisition, the clinical acceptance of US imaging is limited. This further hinders the implementation of extensive screening programs.

To address this challenge, we propose an end-to-end framework for automatic robotic US screening of tubular structures using only real-time US imaging feedback. First, a U-Net is trained for real-time segmentation of vascular structure from cross-sectional US images. Then, we represent the detected vascular structure as a 3D point cloud, which is further used to estimate the centerline of the target structure and its local radius by solving a constrained non-linear optimization problem. Iterating the previous processes, the US probe is automatically aligned to the normal direction of the target structure while the object is constantly maintained in the center of the US view. The real-time segmentation result was evaluated both on a phantom and in-vivo on brachial arteries of volunteers. In addition, the whole process was validated using both simulation and physical phantoms. The mean absolute orientation, centering and radius error (± SD) on a gel phantom were $3.7 \pm 1.6^{\circ}$, $0.2 \pm 0.2\ mm$ and $0.8 \pm 0.4\ mm$, respectively. The results indicate that the method can automatically screen tubular structures with an optimal probe orientation (i.e., normal to the vessel) and accurately estimate the radius of the target structure.

This chapter is organized as follows: In Sec. 5.2, the clinical and methodological background has been summarized. Sec. 5.3 describe the neural network architecture and the implementation details. Then, the details of the proposed close-loop controller is presented in Sec. 5.4. The experimental results on both phantom and in-vivo volunteers are reported in Sec. 5.5. Finally, a detailed discussion and conclusion are presented in Secs. 5.6 and 5.7, individually.

Substantial parts of this chapter have already been published in the following article and are quoted verbatim.

[7] **Zhongliang Jiang***, Zhenyu Li*, Matthias Grimm, Mingchuan Zhou, Marco Esposito, Wolfgang Wein, Walter Stechele, Thomas Wendler, Nassir Navab. "Autonomous Robotic Screening of Tubular Structures based only on Real-time Ultrasound Imaging Feedback," *IEEE Transactions on Industrial Electronics*, Early access, 2021 (*equal contribution).

5.2 Introduction

5.2.1 Clinical Motivation

Peripheral vascular diseases (PVD) are some of the most common diseases, particularly for older people. PVD affects about 20% of adults older than 55 years and an estimated 27 million people in North America and Europe [91]. In the worst case, PVD can lead to critical organ failure, limb amputation, strokes, or heart attacks. The two most common types of vascular diseases are Atherosclerosis and arterial aneurysm, both of which are depicted in Fig. 5.1. Atherosclerosis is a disease in which plaque builds up inside arteries, whereas an aneurysm refers to a weakening of an artery wall resulting in a bulge or distention of the artery. The most common way to diagnose atherosclerosis or arterial aneurysms today is either by measuring the ABI or by inspecting a CTA scan. The drawback of CTA is that the scans expose both the patient and medical staff to ionizing radiation. Conversely, ABI despite being radiation-free, is unable to provide the location of the PVD and is highly user-dependent [93].

B-mode US imaging is a promising alternative to CTA for the diagnosis of vascular diseases due to its lack of radiation [126]. Furthermore, US is widely accessible and cheap, making it popular for real-time diagnosis of internal tissues in clinical practice. Davis *et al.* found that conducting US scans for elderly men can reduce premature death from ruptured abdominal aortic aneurysms by up to 50% [127]. Therefore, regular US examination is important for patients. However, the imaging quality of US is highly dependent on acquisition parameters (i.e., the contact force and probe orientation), which are hard to accurately control in traditional free-hand US, even for experienced sonographers [128]. This leads to a lack of repeatability regarding US acquisition, which negatively impacts the implementation of extensive screening programs.

5.2.2 Robot-Assisted Automatic US Screening

In contrast to conventional free-hand US, automatic screening systems, with stable acquisition quality, can enable regular checks for PVD. Automating the screening procedure by employing a robotic arm can reduce personnel costs and allow sonographers to focus more on analyzing US scans. Furthermore, sonographers can benefit from such a system, as it reduces work-related musculoskeletal disorders.

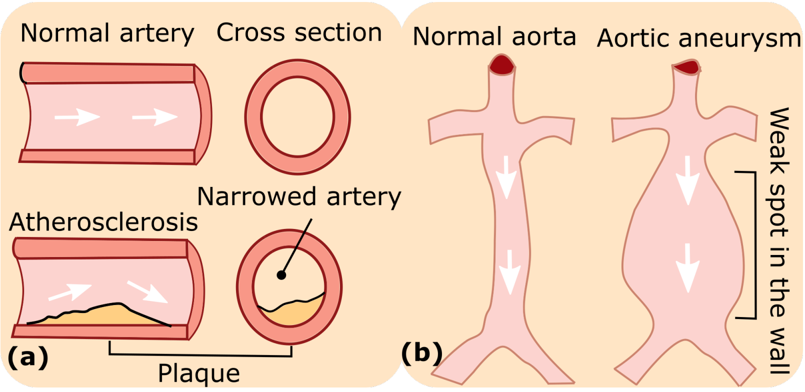


Fig. 5.1. Schematic illustration of the two most common types of vascular diseases: (a) atherosclerosis and (b) arterial aneurysm. *Reprint from [7] with permission of IEEE.*

To develop fully or semi-autonomous RUSS and further obtain enhanced precision in US acquisition, stable and reproducible robotic systems have been widely used in recent studies. Pierrot *et al.* developed a RUSS able to apply a constant force between the probe and contact surface [45]. Gilbertson *et al.* designed a 1-DOF RUSS that reduces image variations by providing a constant contact force [42]. Conti *et al.* proposed a collaborative RUSS that provides an interface for 3D volumetric representation obtained with constant force [129]. In addition to contact force, Huang *et al.* adjusted the probe orientation based on a depth camera [56]. To accurately position the probe in the normal direction of an unknown object, Jiang *et al.* utilized real-time US images and force data for a convex probe [11]. Then, they further developed a mechanical force model based method to accurately position different types of probes for repeatable US acquisition [9]. However, to estimate the normal direction, the probe has to be rotated at a given contact position, which limits the use of these approaches when the probe is moved along a certain trajectory, namely a US sweep. In addition, to track an object, Chatelain *et al.* employed a visual servoing method to maintain the target horizontally centered in the US image using a US confidence map [60]. But the method is not time-efficient, which took around $10\ s$ to reduce the centering error to less than $3\ mm$.

RUSS has also been developed to introduce autonomous and accurate health care solutions for vascular diseases. Balter *et al.* developed a 7-DOF robot for venipuncture using 3D near-infrared and US imaging of peripheral blood vessels [103, 130]. Langsch *et al.* proposed a US-based method for catheter tracking and visualization during endovascular aneurysm repair, while the search path is limited to a vessel path computed from a pre-operative MRI image [65]. Virga *et al.* developed an autonomous framework to acquire abdominal 3D US images for abdominal aortic aneurysms [53]. To automatically obtain a suitable trajectory for different patients, the patient surface obtained from an RGB-D camera was registered to a generic MRI-based atlas. This means the system cannot handle unexpected cases, such as when the target vessel is not visible in the US view. In addition, the vessel radius still needs to be manually measured using the 3D US volume.

5.2.3 US Image Segmentation

US is one of the most challenging modalities for automatic segmentation due to acoustic shadow, poor contrast, the presence of speckle noise, and deformation [131]. To automatically segment vessels from cross-sectional US images, Guerrero *et al.* applied an extended Kalman filter and an elliptical vessel model to determine the vessel boundary using an iterative radial gradient algorithm [132]. However, this method requires careful selection of a seed point inside the boundary. Furthermore, tracking between subsequent frames often fails when the transducer is moved rapidly or if the vessel is deformed. To eliminate the need for manual initialization, Smistad *et al.* proposed an automatic artery detection method and used it to initialize vessel tracking [133]. In addition, to robustly track vessels in a more realistic case, Crimi *et al.* employed the Lucas-Kanade method to follow vessels during sudden lateral transducer movements, even when the vessel is deformed [134]. Nevertheless, the employed ellipse template matching process is time-consuming, and therefore, unsuitable for real-time applications.

Besides elliptical models, adaptive active contours were developed to achieve good segmentation results for clips when variations between US images are taken into consideration [135]. However, this technique requires manual segmentation of the first frame. In addition, Hessian matrix based methods, such as the Frangi filter [136], have been developed to extract tubular structures. But this method cannot accurately extract vessel outlines for further planning US screening.

Recently, machine learning techniques are considered promising alternatives to the classical US image segmentation algorithms. CNN achieved phenomenal success in various computer vision tasks and have been successfully applied to CT and MRI segmentation tasks [137]. However, their use in US image segmentation is limited due to acoustic shadow, poor contrast, and the presence of speckle [131]. Ronneberger *et al.* proposed U-net architecture for biomedical image segmentation [108]. The network is extended from a fully CNN and is considered an effective method for medical image segmentation. Mishra *et al.* developed a fully CNN with attentions for automatic and accurate segmentation of US images [131]. Chen *et al.* employed a recurrent fully CNN architecture to capture salient image features and motion signatures at multiple resolution scales [138]. They reported comparable performance to a human operator for vessel segmentation.

5.2.4 Proposed Approach

In order to online optimize the probe orientation and position and estimate the vessel geometry in real-time, a neural network was employed to provide robust and accurate vascular segmentation results as feedback for a closed-loop controller. The most recent segmentation results were used to characterize the local vessel outline in 3D. Based on the 3D data, the vessel centerline and the radius can be estimated by solving an optimization problem. Based on the estimated vessel centerline, a RUSS can automatically screen blood vessels with no need for pre-operative images. To validate the performance of the segmentation network on in-vivo human tissues, the brachial artery of the forearm was used as the target object in in-vivo tests. Besides, to ensure the safety of involved objects, a simulator with known vessel boundaries was built to theoretically validate the proposed control framework. After fully testing the controller with varied initial settings by simulation, the end-to-end workflow for automatic US scanning (Fig. 5.2) was entirely validated on a physical gel phantom[1].

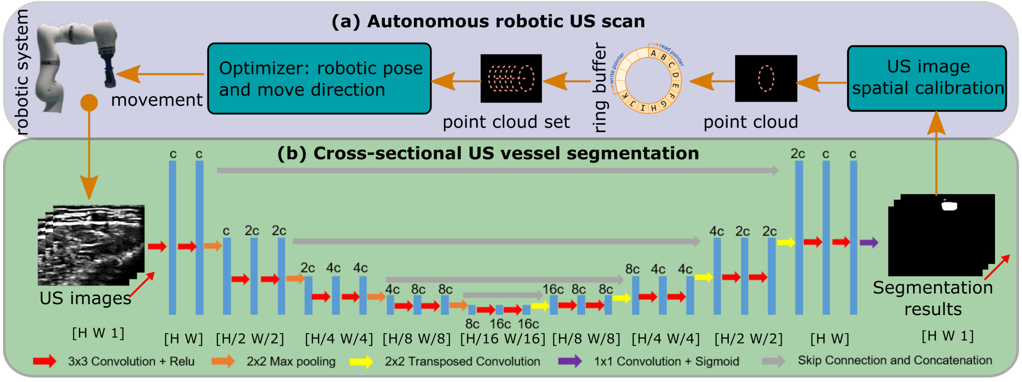


Fig. 5.2. Autonomous image-guided robotic US vessel screening. (a) automatically moving along the target vessel based on real-time segmented results. (b) extracting vessels from cross-sectional US scans using U-Net.The letter“c" represents the number of image channels and “H" and “W” represent the height and width of images, respectively. *Reprint from [7] with permission of IEEE.*

5.2.5 Contributions

In this work, we proposed an end-to-end workflow for autonomous robotic screening of tubular structures based only on real-time US images. This manner was inspired by the way sonographers conduct US scans. First, the US probe is roughly placed on the target tissue. Then, the system can automatically screen along the targeted segment of blood vessels with an optimized pose and estimate the radius of the vessel in real-time. To achieve this, a neural network was trained to segment and track the tubular structures from cross-sectional US images during the screening process. To the best of our knowledge, this work is the first method able to automatically perform a US scan of a vascular structure with optimized probe position and orientation in real-time. The main contributions of this work can be summarized as follows:

- An end-to-end workflow to automatically perform US scans of tubular structures using real-time US images.
- Automatic tuning of the probe orientation and position such that the probe aligns with the normal direction of the structure of interest ($3.7 \pm 1.6^\circ$ in $2.6 \pm 1.5\ s$) and the structure is centered in the US view ($0.2 \pm 0.2\ mm$ in $2.0 \pm 1.0\ s$) during sweeps.
- Starting from a random initial probe orientation within 45° from the optimal orientation, the radius of the tubular tissue was accurately ($0.77 \pm 0.4\ mm$) estimated in a short time ($1.7 \pm 0.5\ s$).

---

[1] The video: https://www.youtube.com/watch?v=VAaNZL0I5ik

Finally, the vascular segmentation method was validated on phantoms and volunteers' arms (brachial artery). The entire screening process was validated both by simulation and on a physical gel phantom.

5.3 Segmentation and Tracking Algorithm

To robustly and accurately segment tubular tissue in real-time from cross-sectional US images, U-Net architecture was employed, consisting of an encoder and decoder. The encoder has a pyramid structure, which is commonly used for segmentation tasks. The decoder contains up-sampling layers that enable the generation of a pixel-wise segmentation mask. Skip connections between encoder and decoder provide the decoder with more image features from the shallow layers, which is beneficial for segmenting detailed structures. The structure of the U-Net is described in Fig. 5.2 (b).

The batch size was set to ten due to GPU memory constraints. This small batch size also improves the generalizability. However, such a small batch size interferes with the batch normalization layers used in the original U-Net architecture. Thus, group normalization [139] was employed in this work.

The U-Net has been proven as a promising method to segment the vessels from US images [99]. However, to train a generic model with good performance for both phantom and human data, a large dataset has to be prepared as the textures of phantom and human images differ. Thus, two models with the same network architecture were trained separately for phantom and human images, respectively. The model developed for an in-vivo human blood vessel (brachial artery) was trained using $1,219$ US images (256×256 pixel) acquired from three volunteers (BMI: 23.2 ± 0.5, age: 28 ± 2), and the model developed for vascular phantoms was trained based on $3,262$ US images (256×256 pixel). In-vivo images of brachial arteries were recorded of the anterior view of the forearm starting below the elbow and towards the wrist with a total distance of $60\ mm$. At least two sweeps were recorded for each subject. To ensure class balance, $1,219$ images were averagely taken from different subjects (around 400 images for each subject). The phantom images were recorded from two custom-designed vascular phantoms with different vessel radii. The larger radius ($7.5\ mm$) was used to mimic big vessels, like the carotid and aorta, while the smaller one ($4\ mm$) was used to mimic small vessels, like the brachial artery. To create random textures in the phantom US images, paper pulp was mixed into the melted gelatin liquid. Then, multiple US sweeps were performed along the vessels with random and varying probe orientation and contact force. Finally, $3,262$ US images were recorded from the two vascular phantoms.

Since US images are sensitive to acquisition parameters and the optimal parameters vary by patient and phantom, it is necessary to provide a robust segmentation for images obtained using different acquisition parameters. To address this issue, the training images were recorded using a variety of parameters (depth, focus, and brightness). Based on the preset file for arterial scans from the manufacturer, the dynamic range and frequency were set to $88\ dB$ and $7.6\ MHz$, respectively. To enable generalizability of the trained models, these parameters were randomly changed in $[70, 95]\ dB$ and $[6.0, 8.5]\ MHz$, respectively, to enhance the diversity of the data sets. Both data sets were carefully annotated with the assistance of a clinician. The dice coefficient is a popular metric used to measure the similarity between the segmentation result and labeled images. Based on this metric, the loss function of the U-Net was designed as follows:

$$L_d = 1 - C_d \tag{5.1}$$

where $C_d = \frac{2\,|G \cap S|}{|G|+|S|}$ is the dice coefficient. In C_d, G are the labeled images, in which the object is carefully annotated, and S are the output images of the network predicting the target object position in US images.

5.4 Force-Compliant Robotic US Screening

This Section describes the method used to optimize the vessel centerline and robotic pose. The optimization was purely based on the most recent US images, without the requirement of additional devices or pre-operative images.

5.4.1 System Calibration

In order to use the US image to control the position of the robotic manipulator, the matrix ${}^{b}_{I}\mathbf{T}$ transforming pixel position from a US image $\{I\}$ into the robotic base frame $\{b\}$ should be calculated. The involved frames are depicted in Fig. 5.3. Thus, ${}^{b}_{I}\mathbf{T}$ can be calculated as follows:

$${}^{b}_{I}\mathbf{T} = {}^{b}_{f}\mathbf{T}\,{}^{f}_{p}\mathbf{T}\,{}^{p}_{I}\mathbf{T} \tag{5.2}$$

where ${}^{j}_{i}\mathbf{T}$ is the transformation matrix used to transfer the position from frame $\{i\}$ to frame $\{j\}$. Frames $\{p\}$ and $\{f\}$ are the probe tip and robot flange frames, respectively.

Since the kinematic model of the robot was given, the transformation ${}^{b}_{f}\mathbf{T}$ from robotic flange to robotic base frame can be obtained using the software API provided by the manufacturer. As for ${}^{f}_{p}\mathbf{T}$, it will be fixed once the probe is attached to the flange. To simplify the calculation and avoid error caused by data truncation, the probe was installed parallel to the frame $\{f\}$ (${}^{b}_{f}\mathbf{R} = (1,0,0;0,1,0;0,0,1)^T$ or $(-1,0,0;0,-1,0;0,0,1)^T$). In addition, the frame origin was set to the central point of the probe tip (Fig. 5.3). The translational part of ${}^{f}_{p}\mathbf{T}$ was obtained from the 3D model.

As for the linear probe, the US elements are physically distributed on the tip within a specific length L_p, which can be obtained from the specification provided by the manufacturer. Based on the given imaging depth D_I and L_p, the pixel position (u, v) can be mapped in physical length $({}^{I}x, {}^{I}y) = (\frac{L_p}{H}u, \frac{D_I}{W}v)$, where H and W are the height and width (in terms of pixel) of the B-mode images, respectively. The origin of frame $\{I\}$ was set at the upper left of the B-mode image, while the origin of frame $\{p\}$ was set in the middle of the probe tip (Fig. 5.3). Thus, ${}^{p}_{I}\mathbf{T}$ is calculated as follows:

$$
{}^{p}_{I}\mathbf{T} = \begin{bmatrix} \frac{L_p}{H} & 0 & 0 & -\frac{L_p}{2} \\ 0 & 0 & -1 & 0 \\ 0 & \frac{D_I}{W} & 0 & \varepsilon_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5.3}
$$

where $L_p = 37.5\ mm$, $D_I = 40\ mm$, and ε_0 is used to represent a small distance from probe frame origin to image frame origin determined by the US element configuration.

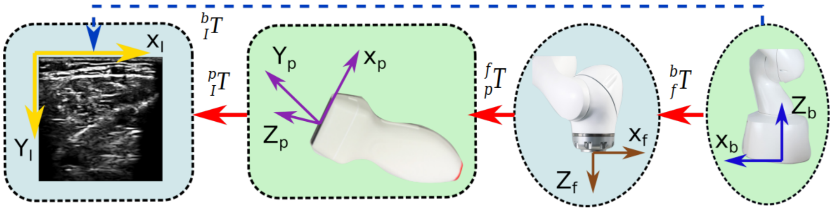


Fig. 5.3. Diagram of involved coordinate systems. *Reprint from [7] with permission of IEEE.*

5.4.2 Impedance Control Architecture

Safety is the most important issue for the development of automatic robotic systems, especially for medical robots. In order to avoid any damage to patients, an impedance controller using built-in joint torque sensors was employed [48]. The involved Cartesian compliant control law has been defined as

$$
\tau = \mathbf{J}^T[\mathbf{F}_d + \mathbf{K}_m e + \mathbf{D}\dot{e} + \mathbf{M}\ddot{e}] \tag{5.4}
$$

where $\tau \in \mathbf{R}^{7\times1}$ is the computed target torque for all joints, $J^T \in \mathbf{R}^{7\times6}$ is the transposed Jacobian matrix, $e \in \mathbf{R}^{6\times1} = (x_d - x_c)$ is the pose error (position and orientation) between the current pose x_c and the target pose x_d in Cartesian space, $\mathbf{F}_d \in \mathbf{R}^{6\times1}$ is the supposed exerted force/torque at endeffector, $\mathbf{K}_m \in \mathbf{R}^{6\times6}$, $\mathbf{D} \in \mathbf{R}^{6\times6}$ and $\mathbf{M} \in \mathbf{R}^{6\times6}$ are diagonal matrices of stiffness, damping and inertia terms in 6 DOF, respectively.

The impedance controller works like a spring with a given stiffness $\mathbf{K}_m$. If an unexpected obstacle is present between the current position and the target position, such as a certain body part of the patient, the robot will stop at the contact position with a certain force. However, if the resistance force is not sufficiently large, the manipulator will overcome the resistance force and continue moving towards the target position. According to [48], stiffness in the direction of the probe centerline is usually set in the range $[125, 500]\ N/m$ for human tissues. Thus, $\mathbf{K}_m$ was set to diag$(1000\ N/m, 1000\ N/m, 300\ N/m, 2\ Nm/rad,\ 20\ Nm/rad, 20\ Nm/rad)$ representing the translational and rotational stiffness in X, Y, and Z and Z, Y. and X direction, respectively. The stiffness in the X and Y directions were empirically set to $1000\ N/m$ to balance the positioning accuracy and compliant propriety in these directions and avoid hard collision. In addition, the damping ratio in all 6 DOFs was set to 0.8. To determine a safe force for patients, the target force F_d of the controller was gradually increased on a volunteer's arm.

Finally, 25 N was set as a software restriction to limit the force exerted by the robot. If contact force $F_c > 25\ N$, the robot will automatically stop to avoid an excessive force.

5.4.3 Estimation of the Vascular Structure Centerline

To build a closed loop system, the robotic movements were optimized based on the received US images as shown in Fig. 5.2. To make the movements smooth and robust, a ring buffer was used to save the newest N_R point clouds generated by transforming the segmentation results from the neural network into the coordinate frame of the robot. Since the cross-section of human blood vessels can be modeled as ellipses [132, 133, 134], the extracted point clouds depicting the vessel boundary can be seen as an elliptic cylindrical surface. Then the vector of vessel centerline ($\vec{n}_v = (n_1, n_2, 1)$) and the vessel radius (r_v) are estimated by solving the following optimization problem:

$$\begin{aligned}
&\min_{\vec{n}_v, r_v, \varepsilon} \underbrace{\frac{1}{2N}\sum_{i=1}^{N}\left(\left(\frac{\|\overrightarrow{CP_i}\times\vec{n}_v\|}{\|\vec{n}_v\|} - r_v\right)^2 + \varepsilon^2\right)}_{\text{optimization term}} \\
&+ \underbrace{\frac{\lambda_1}{2}\left(\arctan\frac{n_2}{n_1} - \arctan\frac{n_2^{'}}{n_1^{'}}\right)^2}_{\text{orientation stabilization term}} + \underbrace{\frac{\lambda_2}{2}(r_v - r_v^{'})^2}_{\text{radius stabilization term}} \qquad (5.5) \\
&s.t. \begin{cases} \varepsilon > 0 \\ r_v > r_l \\ r_v \leq \varepsilon + r_h \end{cases}
\end{aligned}$$

where N is the number of points in the point cloud set, $\mathbf{P}_i = (x_i, y_i, z_i)$ is the point located on the detected vessel boundary, $\mathbf{C} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{P}_i$ is the centroid of the point cloud set. The cross product item represents the distance from $\mathbf{P}_i$ to the estimated centerline $\vec{n}_v$. r_l and r_h are the lower and higher bound of the vessel radius, $r_v \in (r_l, r_h)$, ε is used to soften the hard constraint of the higher bound, to adapt to a serious arterial aneurysm case, wherein the vessel radius becomes much larger than normal. $n_1^{'}$, $n_2^{'}$ and $r_v^{'}$ represent the last optimized results. The latter two terms are used to stabilize the motion of the manipulator. In order to minimize the loss function (Eq. (5.5)), a large pose or radius deviation from the current pose or radius has to be avoided during the optimization process. λ_1 and λ_2 are hyper parameters for tuning the performance of the stabilization terms. Both were set to one in this work.

Considering that the vessel geometry in 2D images is affected by the probe orientation, if the probe is titled form the normal direction, the radius in the 2D image r_v^t (Fig. 5.4 (c)) will be larger than the real r_v (Fig. 5.4 (d)). To accurately estimate r_v in all cases, the newest N_R 3D point clouds are saved in a ring buffer and used to characterize vessel geometry. Then r_v can be approximated by $\frac{1}{N}\sum_{i=1}^{N}\frac{\|\overrightarrow{CP_i}\times\vec{n}_v\|}{\|\vec{n}_v\|}$.

The constrained non-linear optimization problem Eq. (5.5) was solved by implementing a sequential quadratic programming (SQP) optimizer using the NLopt library. Due to the cross product operation, the gradients of the objective function with respect to n_1 and n_2 were computed using the Symbolic Math Toolbox in MATLAB R2020 (MathWorks, USA). However, such an optimizer can get stuck in a local minimum. In our case, the optimizer sometimes yielded a resulting direction vector pointing in the radial direction of the cylinder. In order to circumvent this issue, the original optimization problem was implemented in an iterative "Tick-Tock" manner. The radius r_v was first fixed to a given value, and only the direction vector $\vec{n}_v$ was optimized. Then, r_v was optimized separately with a fixed direction vector. These two steps are executed once in each iteration so that the result will converge to an acceptable local minimum, or even to the global minimum.

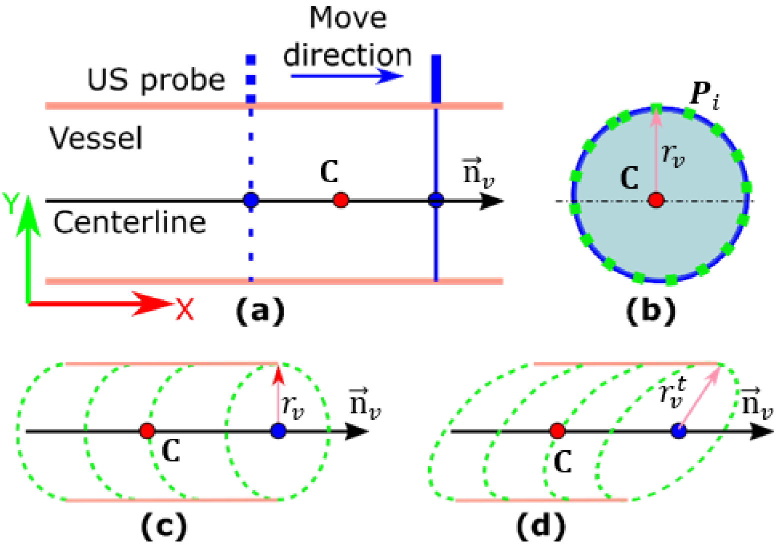


Fig. 5.4. (a) Front and (b) cross-sectional views of blood vessel acquired when the probe is placed normal to the blood vessel. (c) and (d) are the detected point clouds for the sweep when the probe was placed in normal direction and titled direction of vessel centerline, respectively. *Reprint from [7] with permission of IEEE.*

Since the frame rate of the US images is high ($50\ fps$), the detected point clouds in the ring buffer may heavily overlap. This will hinder the optimization process, particularly for real cases wherein the vascular segmentation results are not as perfect as those in a simulation. Considering an extreme case when the probe is paused at a certain position, the detected vessel boundary point clouds in the ring buffer will be distributed on a plane. In such a case, the optimization process could be stuck at a local optimal solution of $\vec{n}_v$ at $\overrightarrow{CP_i}$ (Fig. 5.4). Reducing the overlap will help the optimizer converge to the correct solution (centerline line $\vec{n}_v$ and estimated mean radius r_v of the vessel). To avoid sub-optimal results, the distance between the point clouds Ω_j and Ω_1 in the ring buffer were scaled as follows:

$$\mathbf{P}_i^{'} = \mathbf{P}_i + \mu(\mathbf{C}_j - \mathbf{C}_1) \qquad \forall\ P_i \in \Omega_j \tag{5.6}$$

where $\mathbf{P}_i$ is a point in j-th point cloud Ω_j, $\mathbf{C}_j$ and $\mathbf{C}_1$ are the centroids of the j-th and the first point cloud, respectively, μ is a constant coefficient to move the two point clouds away from each other along the line connecting two centroids. $\mu = 5$ in this work.

5.4.4 Determination of Robotic Pose

In order to obtain a high-quality US image, the US probe is usually aligned in the normal direction of the examined object to reduce the number of scattered echoes [11, 55, 56]. However, previous work [11, 56] only considered the normal direction of the skin surface above target tissues. In order to find the normal direction of a vessel tissue, which is located below the skin surface, both camera-based [56] and force-based methods [9, 11] are not applicable.

However, as detailed in the previous Section, the present method can estimate the centerline vector $\vec{n}_v$ of a blood vessel from a set of US images; allowing normal alignment of the US probe to the target vessel by aligning the probe's Y_p axis with $\vec{n}_v$. Once the alignment has been done, the US image will remain normal to the vessel when the probe is rotated around $\vec{n}_v$. However, this rotation will result in non-homogeneous deformation in both sides of the US image or even cause a gap between the US probe and skin. Thus to fully obtain the proper probe orientation, and to achieve the best US image quality, the probe Z_p axis is aligned with the normal of the contact surface $\vec{n}_s$. Since this is evaluated on a gel phantom with a flat surface, $\vec{n}_s$ is approximated using the normal direction of a plane composed of three neighboring points manually selected using the robot. After this procedure is completed, the target probe orientation is determined ($X_p = Y_p \times Z_p$).

In addition, to maintain the target vessel horizontally centered in the US image for a better view, the position of the probe also needs to be adjusted based on US images. Since the relative movement ΔP is continuously updated using real-time US images, the target vessel is able to be maintained around the horizontal center of the US view, even when the displayed tissue is deformed due to the contact of the probe.

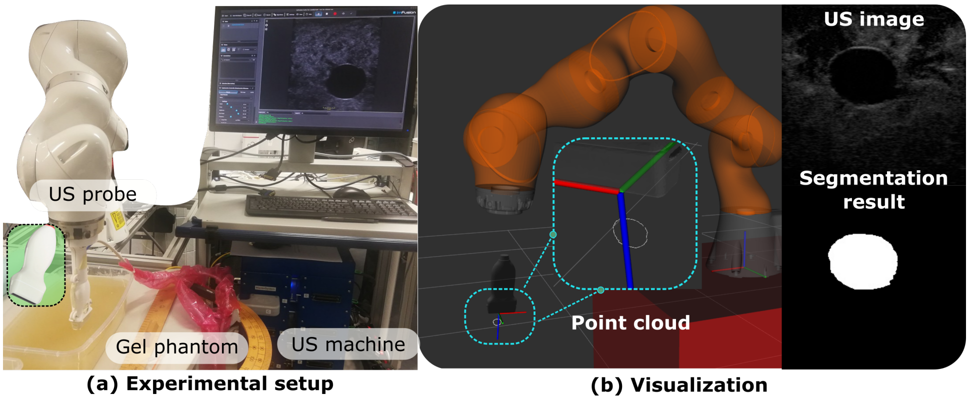
$$\Delta P = {}^{b}_{I}\mathbf{T} \left[\frac{H}{2} - x_c^I,\ 0,\ 0,\ 1 \right]^T \tag{5.7}$$

where $x_c^I = \frac{1}{N} \sum_{i=1}^{N} x_p^i$ is the horizontal center of the detected vessel in the current image, and x_p^i is the horizontal value of the point **P** in the ring buffer.

5.5 Results

5.5.1 Experimental Setup

The overall experimental setup is shown in Fig. 5.5 (a). A linear probe (CPLA12875, Cephasonics, California, USA) was attached to the end-effector of a robotic manipulator (KUKA LBR iiwa 14 R820, KUKA Roboter GmbH, Augsburg, Germany) using a 3D-printed mount. The used robot has accurate built-in torque sensors in all joints. The B-mode US images (50 fps) were obtained using a Cephasonics machine (Cephasonics, California, USA) via USB interface and the images were visualized using ImFusion software (ImFusion GmbH, Munich, Germany). To access real-time images from different computers, the US image stream was published to ROS. In addition, the robot was controlled via iiwa stack developed based on the ROS [48]. The control commands and the robot status were exchanged at $100\ Hz$.



(a) Experimental setup **(b) Visualization**

Fig. 5.5. (a) Physical setup of experiments on a gel phantom. (b) Real-time visualization in RVIZ. The real-time US image and segmentation results were also visualized in RVIZ. *Reprint from [7] with permission of IEEE.*

To build a ballistic gel phantom with non-homogeneous acoustic impedance mimicking human tissues, paper pulp ($3-5\ g/L$) and gelatin powder ($175\ g/L$) were mixed into water. After solidification, two round tubes were used to create holes inside the phantom mimicking vascular structures. Since the radii of the tubes are constant along the centerline, the inner radius of the created holes can also be seen as a constant value, which was approximated using the average radius directly measured from the removed cylinder portion of the phantom. To completely validate the proposed end-to-end automatic US scan approach, one of the phantoms, with a radius of $7.5\ mm$, was used for physical experiments. The marching velocities for both simulation and phantom experiments in the direction $\vec{n}_v$ were set to $1\ cm/s$. To monitor the screening process, the RUSS and the point cloud of the detected vessel were visualized in ros-visualization (RVIZ) as shown in Fig. 5.5.

5.5.2 Segmentation Results on Phantom and Human Arm

The model was optimized using the ADAM optimizer [112] on a single GPU (Nvidia GeForce GTX 1080). The learning rate was set to 0.001 in the beginning and decreased by a factor of ten when the loss changes were less than 0.0001 for ten subsequent steps. The decrease of the learning rate in the latest step of training can help further reduce loss. Besides, the U-Net models were trained on 90% of the images of each data set while the remaining 10% of the images were used for validation. It can be seen from Fig. 5.6 that both training loss and validation loss were quickly reduced at the beginning and gradually converged after $1,000$ iterations. The performance of the trained model on unseen images of brachial arteries and phantoms are summarized in TABLE 5.1.

The trained model can segment the vascular structure in $5.1\ ms$ and the average dice coefficients ($\pm SD$) were 0.93 ± 0.05 and 0.98 ± 0.02 for human tissue (26 images) and phantom (100 images). To validate whether the trained model is able to robustly and accurately segment the target from cross-sectional US images, the phantom was rotated to different orientations relative to the probe as shown in Fig. 5.7 (a) and (b). In addition, deformation is unavoidable

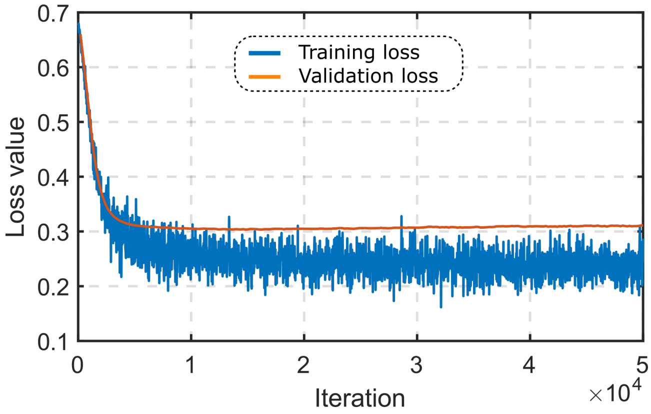


Fig. 5.6. Training loss and validation loss of the segmentation network. *Reprint from [7] with permission of IEEE.*

Tab. 5.1. Performance of Segmentation Algorithm

Dataset	Dice Coefficient	Time (ms)
Phantom	0.98 ± 0.02	5.1
Brachial artery	0.93 ± 0.05	5.1

for US images due to a non-zero contact force. To validate the performance of the proposed segmentation method on deformed US images, the pressure between the probe and phantom was manually increased until the vessel was severely deformed as shown in Fig. 5.7 (d). As can be seen in Fig. 5.7, the trained model is able to accurately and robustly segment the target tissue (mean dice coefficient is over 0.98) from the US images with different extents of deformation. In order to further consider the potential for real clinical use, we tested the performance on arms of volunteers (brachial artery). The network also successfully extracts the target vessel from a continuous sweep in the testing data set as shown in Fig. 5.8. The mean dice coefficient for in-vivo test was 0.93 ± 0.05. This means that the segmentation method can be used on in-vivo brachial artery.

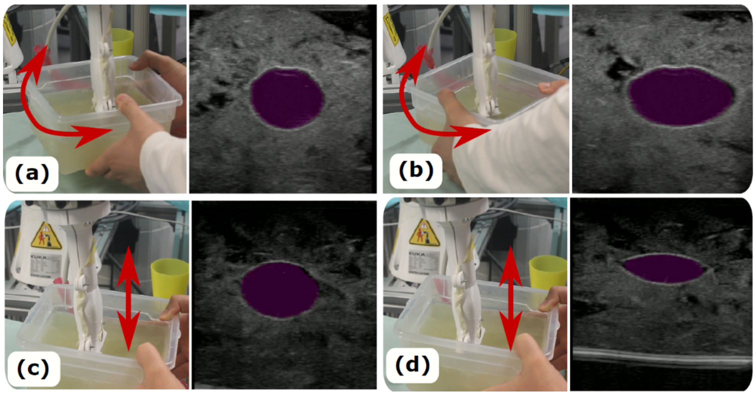


Fig. 5.7. Sample segmentation results on a phantom. *Reprint from [7] with permission of IEEE.*

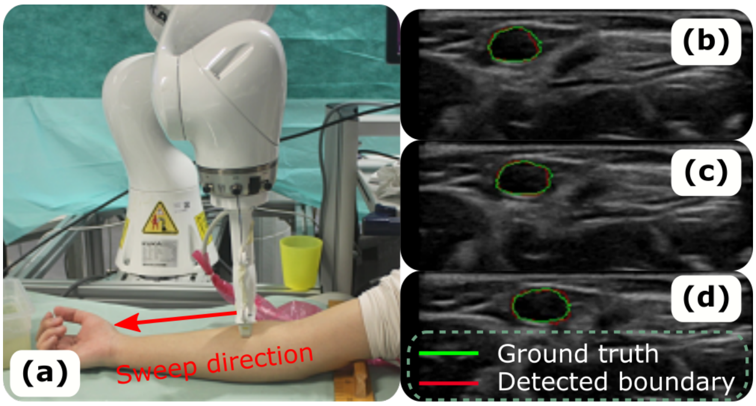


Fig. 5.8. Sample segmentation results on a human. *Reprint from [7] with permission of IEEE.*

5.5.3 Automatic Screening in Simulation

To avoid damage of objects under evaluation, the proposed end-to-end control framework for automatic vessel screening was first validated in a simulator. Here, the elliptical vessel boundary point clouds were generated in MATLAB R2020 using a 2D plane to cut a cylindrical surface mesh. The resulting 2D point clouds were fed into the ring buffer to optimize the vessel radius and the vessel centerline by solving Eq. (5.5). Then, the probe could be gradually aligned with the normal direction of the vessel and moved along the vessel centerline. To demonstrate the performance of the proposed approach by simulation, the absolute radius error e_{ra} between the estimated radius and the preset value, and the absolute orientation error e_{or} between computed and desired poses are shown in Fig. 5.9. Since the radius of some healthy arteries like the aorta can be $16.8\ mm$ for males [140], the radius was set to $10\ mm$ in the simulation.

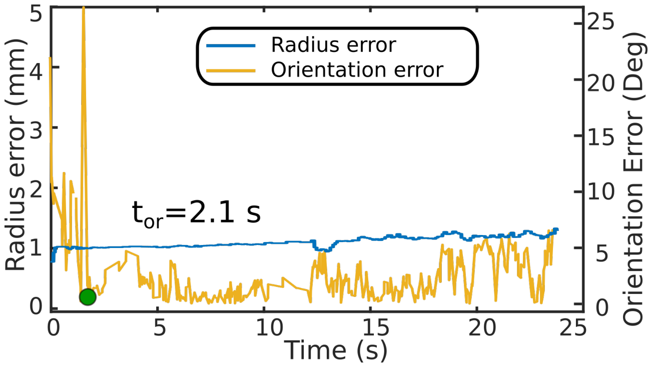


Fig. 5.9. Performance of the optimization method by simulation. *Reprint from [7] with permission of IEEE.*

It can be seen from Fig. 5.9 that the proposed optimization method effectively decreases e_{or} to $5°$ from a random initial position in a short amount of time ($2.07\ s$). The largest angular error ($22.5°$) only occurs at the beginning of the process. This is because heavy overlap of the first few US images results in failure of the optimization method to find a good estimation of the vessel centerline. Once a good estimation of the vessel centerline is found, e_{or} kept within a small value. The absolute mean (± SD) e_{or} over all 330 scans is $2.7 \pm 3.3°$. This error is very close to the previous state-of-the-art approach for accurate positioning of US probes on a gel phantom $2.9 \pm 1.6°$ [9]. More importantly, the proposed method can online optimize the probe pose during sweeps while previous methods with similar accuracy only work at the given position [9, 11].

In addition, the estimated radius quickly converges to a stable value that is close to the given radius of $10\ mm$ very fast showing that the proposed method can effectively estimate the real radius without the strict requirement to align the probe normal to the blood vessel. The absolute mean (± SD) e_{ra} is $1.16 \pm 0.1\ mm$. Compared with the real radius ($10\ mm$) preset in the simulation, the deviation of e_{ra} $0.1\ mm$ is quite small (1%). Therefore, we can conclude that the proposed method demonstrates the ability to stably predict the radius with an error of less than 11.6%.

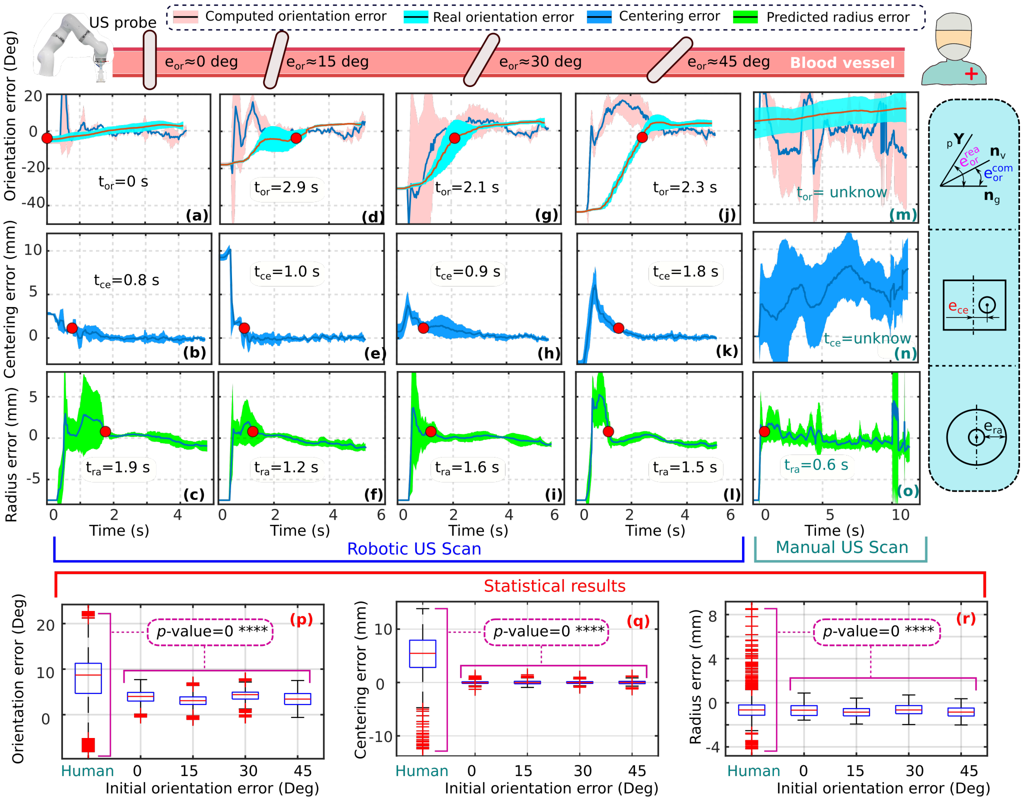


Fig. 5.10. Performance of the proposed automatic and manual screening methods on a gel phantom. The e_{or}^{com}, e_{or}^{rea}, e_{ce} and e_{ra} are intuitively described in the right of the figure and these errors are indicated by the peach, aqua, blue and green plots, respectively. The shadowed area represents $(mean - SD, mean + SD)$ over the repeated experiments, while the curves inside the shadowed areas are the average results. The results for robotic autonomous scans when the probe was placed in different initial orientations are shown in (a)-(i). (m) and (n) depict the e_{or}^{rea} and e_{ce} by blue and red lines, respectively. Based on US sweeps performed by a human operator, the e_{or}^{com} and e_{ra} computed using the proposed optimization method are described in pink and black lines in (m) and (o). (p), (q) and (r) are the statistical results of e_{or}^{rea}, e_{ce} and e_{ra} obtained from the steady periods of all experiments. The p-value is the probability from t-test used to compare the performance of human operators and robotic screening in terms of e_{or}^{rea}, e_{ce} and e_{ra}, respectively. The scan path was about 8 cm. *Reprint from [7] with permission of IEEE.*

5.5.4 Automatic Screening on a Phantom

To further validate the performance of the proposed automatic screening system, experiments were carried out on a physical phantom with different initial orientation errors ($e_{or} \approx$

Tab. 5.2. Absolute Accuracies and Converge Times of Different Methods on phantom (Mean±SD)

Methods and used signals	Orientation		Centering		Radius		Sweep time (s)
	e_{or} (deg)	t_{or} (s)	e_{ce} (mm)	t_{ce} (s)	e_{ra} (mm)	t_{ra} (s)	
Human operator	8.5±4.6	unknown	5.7±3.4	unknown	×		≈13
Camera [52]	7.3±3.1#	≈2	×		×		≈6
Force [5]	2.7±1.9	>30	×		×		×
Force and US image [7]	3.1±1.0	>20	×		×		× work for
US image [56]	<2.0*	>7	<3	>10	×		× given points
Proposed method	3.7±1.6	2.6±1.5	0.2±0.2	2.0±1.0	0.8±0.4	1.7±0.5	≈6

×: method is not appliable. #: Recompute based on RealSense D435. *: Only in-plane orientation adjustment.

$0°, 15°, 30°, 45°$). The screening procedure was repeated ten times for each initial e_{or} setting. The corresponding results ($mean \pm SD$) of the robotic scans and ten manual sweeps results on the same phantom are shown in Fig. 5.10. Unlike the simulation, for the real case, we calculated both the real orientation error e_{or}^{rea} between the ground truth of the centerline $\mathbf{n}_g$ and the current robotic pose (Y axis of frame $\{p\}$) as well as the computed orientation error e_{or}^{com} between the computed centerline $\mathbf{n}_v$ and $\mathbf{n}_g$ (Fig. 5.10). To determine $\mathbf{n}_g$, the probe was manually placed at the start and end points of the tubular tissue. Then, the pixel position of the vessel centroid was transferred to frame $\{b\}$ using the spatial calibration described in Section IV-A. Thus, the vessel centerline in frame $\{b\}$ can be represented by the line connecting these two centroids. In addition, the horizontal centering error e_{ce} was represented by the horizontal distance from the computed vessel's centroid of the current image and the vertical centerline of the US view.

Performance of Robotic Screening

It can be seen in Fig. 5.10 that e_{or}^{com}, e_{or}^{rea}, e_{ra} and e_{ce} for all experiments with different initial e_{or} can converge to zero using the proposed method. As the initial error increases from zero to $45°$, the time t_{or} required to align the probe in the desired direction also increases from zero to $2.9\ s$. In most cases, t_{or} is longer than the time needed to stably estimate the vessel radius t_{ra}. This is because the radius estimator can predict the current radius even when the probe is not placed in an optimal orientation (i.e. normal to the vessel). Since the robot starts moving only when the ring buffer is full, the beginning images are highly overlapped. The overlapped images will result in suboptimal results of $\vec{n}_v$ and e_{ra} as described in Section IV-C, which leads to large e_{ra} and e_{or} at the beginning of the optimization. Then, e_{ra} decreases quickly to a low value near zero ($< 1.8\ s$) for all robotic scans. Regarding the centering error e_{ce}, this value approaches zero ($< 0.5\ mm$) after t_{ce}, which is the consumed time for horizontally centering the target vessel in the US view. Such small error is because the relative movement ΔP (Eq. (5.7)) can be accurately calculated with respect to pixels.

To compare the performance of the proposed method with existing approaches, the absolute e_{or}, e_{ce} and e_{ra} and the times t_{or}, t_{ce} and t_{ra} are shown in TABLE 5.2. The mean e_{or}, e_{ce} and e_{ra} over all 40 experiments with different initial orientations in $[0, 45°]$ were $3.7 \pm 1.6°$, $0.24 \pm 0.19\ mm$ and $0.77 \pm 0.4\ mm$. Regarding probe orientation adjustment, the proposed method is not the most accurate; however, our approach would be the best if the time needed to adjust the orientation is considered ($2.6 \pm 1.5\ s$) as well. The time needed to accurately center the object and predict the radius are $2 \pm 1\ s$ and $1.7 \pm 0.5\ s$, respectively.

Performance of Manual Screening

To compare the performance of the proposed method and of human operators, three experienced volunteers were asked to manually perform ten scans (30 times in total) on the same phantom. The volunteers were required to adjust the probe orientation from a random pose in the normal direction of the vessel centerline and to maintain the optimal pose during the entire sweep. During manual scans, the built-in "hand-guiding" mode was activated, which allows operators to freely move the probe attached to the robotic flange. The probe pose was tracked in real-time based on the robotic kinematic model. One of the results was randomly selected as an example shown in Fig. 5.10 (m) and (n). Additionally, the point clouds generated from human scans were fed to the proposed optimization algorithm to predict the vessel centerline and radius. The estimated e_{or}^{com} and e_{ra} are depicted by pink and black lines, respectively, in Fig. 5.10 (m) and (o).

It can be seen from Fig. 5.10 (m) that e_{or}^{rea} and e_{ce} cannot converge like robotic US scans. The average e_{or}^{rea} continuously increases from 3° at the beginning to around 14° at the end for manual scans. In addition, the deviation of e_{or}^{rea} is constantly around 11°, which cannot be reduced as the scanning progress. Regarding e_{ce}, both the average and the deviation values cannot converge to zero. They constantly varies in the range of $[0, 8\ mm]$ during the manual scan. This is because the human perception's ability is not good at accurately identifying small differences in orientation and position. Since such tasks require good hand-eye coordination, performing as well as RUSS is challenging for human operators, particularly when multiple objects (optimal orientation, position, and movement along vessel) are required simultaneously.

Statistical Comparison between Robotic and Manual Screening

To quantitatively demonstrate the advantages of the proposed approach, statistical results of e_{or}^{rea}, e_{ce} and e_{ra} of manual scans and robotic scans with different initial orientation are shown in Fig. 5.10 (p), (q) and (r), respectively. For the robotic scans, the proposed method can properly adjust the probe orientation and position for centering the object in the US view, even when the initial orientation is up to 45° from the surface normal.The t-test has been employed to compare the performance of human operators and robotic screenings in terms of e_{or}^{rea}, e_{ce} and e_{ra}. The probability values p-value are zero, which means that there is a significant difference ($> 95\%$) between the performance of manual scans and robotic scans. This phenomenon is consistent with the intuitive results witnessed from Fig. 5.10 (p), (q) and (r) that the stable errors for robotic screenings are much smaller than the ones achieved in manual scans. The median e_{or}^{rea} and e_{ce} are 8.7° versus 3.7° and $3.1\ mm$ versus $0.02\ mm$ for manual scans and robotic scans, respectively. Since the centering movement for robotic scans was controlled in terms of pixels, e_{ce} can be very close to zero. Regarding e_{ra}, human operators were not able to measure the vessel radius in real-time during US sweeps. Since the proposed method can estimate the radius based on the most recent frames, the method was used to predict the radius for the manual scans in Fig. 5.10 (r). The results demonstrate that the radius prediction method can also effectively predict the radius. The median e_{ra} for manual scans was $-0.7\ mm$, which is very close to the robotic scans $-0.8\ mm$. But it can also be observed that the maximum values of absolute e_{or}^{rea}, e_{ce} and e_{ra} for manual scans are much larger than the errors for robotic scans (22.6° versus 7.8°, $13\ mm$ versus $1.4\ mm$ and $8.5\ mm$

versus 2 mm). This is due to the manual scans' relative instability compared to the robotic scans (i.e., scan velocity and probe pose).

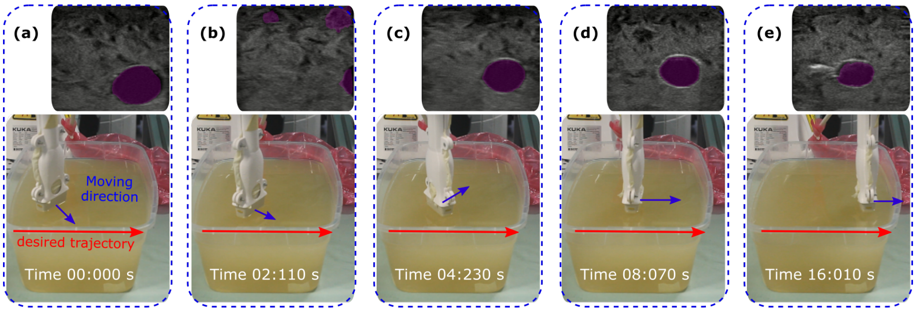


Fig. 5.11. Performance of the optimization method on a phantom with an suboptimal initialization. *Reprint from [7] with permission of IEEE.*

Representative Robotic Screening Procedures

To further show the detailed process of the proposed end-to-end screening system, a full US sweep (16 cm) starting from a large initial $e_{or} \approx 45°$ is shown in Fig. 5.11. The probe reached the desired orientation at 8 s and then, the optimal orientation is maintained until the end of the sweep at 16 s. This shows that the proposed method can perform automatic US screening based on real-time US imaging. In addition, it is noteworthy that false-positive cases of real-time segmentation may occur during scanning (e.g., Fig. 5.11 (b)). To correctly select a good candidate in such a case, the distances between the selected vessel in the previous frame and the detected results in the new frame were calculated. Since the vessel is a continuous object, the detected result with the smallest distance is finally selected to reduce the chance of tracking the wrong object. To analyze whether the proposed method is able to run in real-time, the image processing and optimization overhead were measured. The entire process took 32 ms on average while the U-Net only took 5 ms for image processing (TABLE 5.1). Therefore, the system can be run in real-time.

5.6 Discussion

This work enables RUSS to automatically perform US sweeps for tubular structures without any requirement for pre-operative images. The proposed approach is able to optimize the probe pose (position and orientation) and calculate vessel radius in real-time during US sweeps. Since autonomous RUSS overcomes the limitation of human hand-eye coordination, a stable and accurate prediction of the radius can be achieved. Significantly, due to the use of multiple frames, the proposed optimization approach is able to accurately calculate the vessel radius when the US probe is titled from normal direction. Compared with existing methods, the approach also demonstrated the advantage of optimizing multiple objects meanwhile. The errors of the probe orientation, target centering and estimated radius are mainly caused by the error of the image segmentation results. In addition, the difference between the real orientation error and estimated orientation error is because of the compliant behaviour of the controller. To guarantee patients' safety and good contact condition between probe and object surface, the impedance control was employed, however, this controller allows positional error to achieve compliant performance.

In addition, the main limitation of this work is that the segmentation network was only tested on the brachial artery in the forearm. A more generic model for segmenting different arteries has to be developed to further test the proposed approach in different clinical trials. Considering variations between patients, to make a network accurately segment target tissues for most people, Prevost *et al.* suggested that several tens of US sweeps with different acquisition parameters on several tens of volunteers are required to be used for training [99]. Alternatively, to overcome the difficulties in collecting a larger number of in-vivo data sets, large-scale simulated US data and small in-vivo data sets can be used together to train an generic segmentation network, as in [141]. Beside, the technique for correcting the pressure-induced deformation in US images [6] can be further integrated to generate accurate 3D anatomy.

5.7 Conclusion

In this work, we introduced an end-to-end workflow for autonomous robotic screening of tubular structures based only on real-time US images. The proposed method can automatically adjust the probe position to horizontally center a tracked object and tune the probe orientation to the normal direction of the target vessel during US sweeps. Additionally, the radius of the target object is calculated in real-time, even when the probe is titled from the normal direction. The results demonstrated that the proposed method is able to automatically perform US sweeps and accurately estimate the geometry of the target vessels both in simulation (e_{ra}: $1.16 \pm 0.1\ mm$, e_{or}: $2.7 \pm 3.3°$) and for a real phantom (e_{ra}: $0.77 \pm 0.4\ mm$, e_{or}^{rea}: $3.7 \pm 1.6°$, e_{ce}: $0.24 \pm 0.19\ mm$). The development of autonomous RUSS is a promising way to overcome inter-operator variability providing repeatable US images. In addition, with such a RUSS, sonographers can focus on diagnosis to fully utilize their experience and knowledge. This approach could be integrated with autonomous diagnosis techniques to further pave the way for a fully automatic US-guided intervention system [142].

6

Learning US Examination from Expert Demonstrations

6.1 Overview

This chapter presents the contributions of this thesis concerning learning-based autonomous US scans. Due to the missing of the key ability to learn US scans from expert demonstrations, most robotic US imaging systems can only assist sonographers by maintaining contact force; or semi-/fully automatically performing a predefined scanning task.

To this end, we present an intelligent robotic sonographer, which can autonomously “explore" target anatomies and navigate a US probe to a relevant 2D plane by learning from expert’s demonstrations. To intuitively transfer high-level physiological knowledge, a number of demons starting from random positions and ending at the target objects are performed by experts. The clinician’s intention is extracted by a reward function computed using ranked pair-wise image comparisons generated from all demos. To avoid inconsistent objectives, all demons are ranked and cleaned using Gaussian distribution. The inferred rewards of individual B-mode images are then used to locate target anatomies by successively performing coarse and fine localization. To validate the performance of the proposed approach, two representative experiments were performed on a vascular phantom and two ex-vivo animal organ phantoms (chicken heart and lamb kidney). The results show that the method is able to automatically navigate a US probe to standard scan planes.

This chapter is organized as follows: In Sec. 6.2, the clinical and methodological background has been summarized. Then, Sec. 6.3 presents the proposed advanced approach enables learn the latent reward from suboptimal demonstrations from clinical experts. Afterwards, the autonomous navigation approach guiding probe to a target standard plane is presented in Sec. 6.4. Finally, the experimental details and a through conclusion are presented in Secs. 6.4.2 and 6.6, individually.

6.2 Introduction

6.2.1 Clinical Motivation

Ultrasound is one of the most widely used non-invasive imaging techniques to visualize internal anatomies. Unlike CT and MRI, US imaging is real-time, low cost and radiation free [19]. Conventional US acquisition requires sonographers to manually orient a US probe in specified poses called standard scan planes. Such planes are often used to perform quantitative biometric measurements [143, 144], image-guided interventions [138, 145] and to identify abnormalities, as shown in longitudinal view of organs (i.e., blood vessel, kidney) in Fig. 6.1. However, substantial experience and visuo-tactile skills are required to identify and accurately align a probe with standard planes. The potential for variation between different operators hinders consistent and repeatable diagnosis, particularly for inexperienced sonographers. Therefore, a robotic US system (RUSS) with the ability to discover standard planes could bring advantages by reducing the workload of sonographers and obtaining standardized and operator-independent US images.

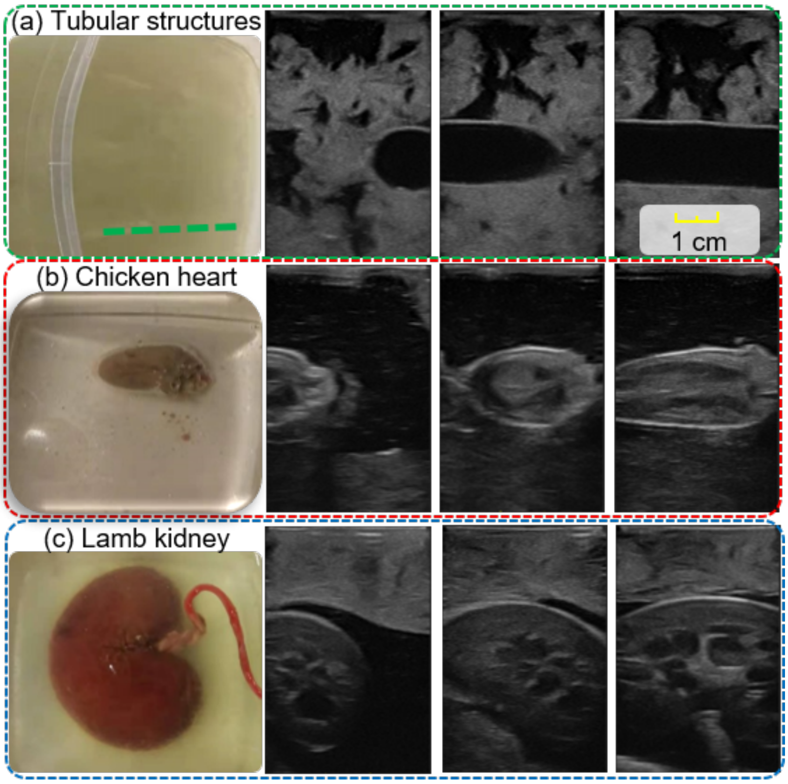


Fig. 6.1. Illustrations of standard planes for different organs. (a) the longitudinal view of a mimicked vesicular phantom, (b) longitudinal view of a chicken heart, and (c) longitudinal view of a lamb kidney. From the 2nd column to the 4th column are B-mode images of different objects recorded by a human operator manually aligning the probe pose towards standard planes. US imaging depth is $5\ cm$.

6.2.2 Robotic US Imaging System

Due to the advantages of accuracy, stability, and repeatability, robotic technology has been employed to develop both a fully or a semi-autonomous RUSS. Pierrot *et al.* developed a RUSS with the ability to maintain a constant force between the probe and subject's skin for cardiovascular disease prevention [45]. To further quantify US acquisition for stable and repeatable US images, Jiang *et al.* proposed a method to automatically position a linear probe to the normal direction of the object skin using a confidence map and the estimated Cartesian force at TCP [11]. Then, a mechanical force model-based method was proposed to optimize the orientation of both convex and linear probes [9].

Regarding automatic robotic US scanning, Virga *et al.* proposed an approach to acquire 3D US images for abdominal aortic aneurysms screening [53]. The scan path was determined based on the registration between the patient's surface and a generic MRI-based atlas. To eliminate the requirement for the pre-operative images, Huang *et al.* proposed a method to automatically determine the scan path based on the extracted skin surface using a depth camera [56]. To fully utilize human experience, Abolmaesumi *et al.* develop a shared controller to visualize carotid arteries in 3D [44]. The aforementioned RUSSs are mainly developed to automatically and accurately visualize the anatomies of interest. However, these systems are not designed for automatically discovering and navigating a probe to standard scan planes, which are important for quantitative measurement [143, 144] or image-guided interventions [138] in clinical practice.

6.2.3 Detection and Navigation of Standard Planes

Due to the potential deformation of soft tissue and the often hard-to-interpret US imaging, guiding a probe to correct planes is a highly sophisticated task, which requires years of training [146]. Furthermore, limited by hand-eye cooperation, these tasks suffer from low reproducibility and large variation between operators [22]. These drawbacks severely affect the clinical acceptance of the US modality towards repeatable, quantitative, and accurate US imaging. To address this challenge, Chen *et al.* proposed a learning-based approach to locate the fetal abdominal standard plane in US videos using a deep convolutional neural network (CNN) [147]. Each input video frame was processed by a classifier to detect the standard plane using a sliding window. However, this approach requires carefully annotated images for training and cannot be run in real-time. To bridge the gap to real examination, Baumgartner *et al.* proposed the SonoNet to detect standard views from US images in real-time [144]. This approach achieved good performance for data acquired in freehand fashion, but it cannot navigate an probe towards target imaging planes.

To effectively navigate a probe to standard planes, Droste *et al.* trained a policy network to predict the final goal pose and the next movement action based on the recorded US images and corresponding probe poses from expert demonstrations [148]. The network uses a gated recurrent unit (GRU) and was trained based on paired US images and corresponding probe orientations recorded from experts' demonstrations. Due to the limitation of the inertial measurement unit (IMU), only rotation movement was considered. In addition, to successfully predict the next movement, the relative movement from the previous frame to the current frame has to be properly estimated, which hinders the possibility to apply such a method in RUSS due to the prediction error accumulation.

To guide a probe in a real scenario, Hase *et al.* employed a deep Q-network (DQN) to navigate the probe towards sacrum in a 2D grid word; and the navigation process is terminated by feeding the real-time US images into a binary classifier [149]. Since rotation was not considered, a good initialization of orientation has to be given. In addition, the binary classifier limits the accuracy of the result because the search process will stop once the classifier is triggered (reach the preset threshold rather than the optimal value). To robustly search and align the probe to the paramedian sagittal oblique plane (a standard plane used in spine US examination), Li *et al.* considered both rotational and translational movements based on real-time images obtained from a simulator using a deep Reinforcement Learning (RL) framework [150]. The results achieved in unseen data was only 46% successful, which is much smaller than the reported result for training data, of 92%. Besides, this approach was only evaluated in simulation with a virtual probe. How to bridge the gap between the simulation and real world scenarios is still an open challenge, particularly for clinical applications where safety is critical.

6.2.4 Learning from Demonstration

Since RL algorithms typically require a lot of interaction with the environment to successfully learn policies to solve a specific task [151], most of the aforementioned studies were trained in simulation or using recorded US examination videos. Besides RL methods, some works tried to directly teach RUSS to perform US scans based on expert demonstrations, which can effectively alleviate the complexity of robotic programming. In addition, such a system can take full advantage of senior operators' experience and help train young operators. Learning from demonstration (LfD) can be broadly categorized into two approaches: imitation learning (IL) [152] and inverse reinforcement learning (IRL) [153]. Imitation learning directly learns a predictive model to estimate the next action based on the current state, which requires optimal demonstrations because the fundamental logic of such an approach is to imitate the behavior rather than understand the latent objective. On the other hand, IRL aims to represent a given task by a reward function with respect to state features. Common approaches assume that the reward is expressed as a linear combination with the feature counts [153] or that the likelihood of observing a demonstration is exponentially higher for higher rewards than lower rewards [154]. Due to acoustic shadows, poor contrast, speckle noise, and potential deformation in resulting imaging, guiding a probe to correct planes is a highly sophisticated task in clinical applications [131]. This means that expert demonstrations are frequently and inherently sub-optimal and even contradictory. Therefore, the popular maximum-entropy IRL method [155] is not suitable for our task.

In the field of robotic US, Mylonas *et al.* employed Gaussian Mixture Modeling (GMM) to model the demonstration in a probabilistic manner [156], but only took the trajectory into consideration. Since the resulting US images were not considered, all the demonstrations had to be started from the same point, which hinders the potential to apply the trained model to different setups. To achieve good performance, this approach has strict requirements on the initialized position and phantom position. Burke *et al.* introduced a probabilistic temporal ranking model which assumes that the images shown in the later stage are more important than the earlier images [157], allowing for reward inference from sub-optimal scanning demonstrations. They use this model to coarsely navigate a US probe to a mimicked tumor inside of a gel phantom, using an exploratory Bayesian optimisation policy to search for scanning positions that capture images with high rewards. However, in real scenarios, patient safety is critical, and it is not realistic to richly interact with patients to gain experience to find an optimal viewpoint.

6.2.5 Proposed Approach

To assist the sonographer in achieving consistent and repeatable diagnosis during US examinations, we developed an intelligent RUSS to navigate the probe to standard scan planes by learning from few demonstrations. To the best of our knowledge, this is the first work that aims to learn the underlying skill and anatomical knowledge directly from human operators to align a probe to standard scan planes automatically. The ability to learn from demonstrations allows sonographers to intuitively transfer their knowledge to a RUSS without the requirement of any robotic programming. This is still a challenging task in the area of robotics skill learning, particularly when the demonstrations are not optimal [158]. Inspired by [157], a global probabilistic spatial ranking (GPSR) method was developed to learn the latent skills and overcome the potentially biased results caused by inconsistent demonstrations. The GPSR mainly consists of a convolutional variational autoencoder (VAE) [159] used to extract the features of US images and a fully connected network (FCN) used to predict the reward associated with individual frames. Considering the time efficiency and potential movement of the patient during the procedure to obtain optimal policy navigating the probe to the target plane, it is impractical to train an agent using RL [149, 150] or apply Gaussian process path planning [157] in a real world scenario. Therefore, we divided the navigation process into coarse and fine tuning procedures. Regarding the first step, a US sweep over the target anatomies was performed, and then a virtual probe was generated on the upper surface of the computed 3D volume. By assigning various probe poses, a larger number of simulated 2D images were created. Based on the predicted reward for individual synthetic images, a greedy algorithm was employed to detect the desired plane. Finally, to further bridge the gap between simulated images and real images, the probe was finely adjusted around the coarse location to achieve a more precise result, like human operators. An overview of the proposed approach is shown in Fig. 6.2.

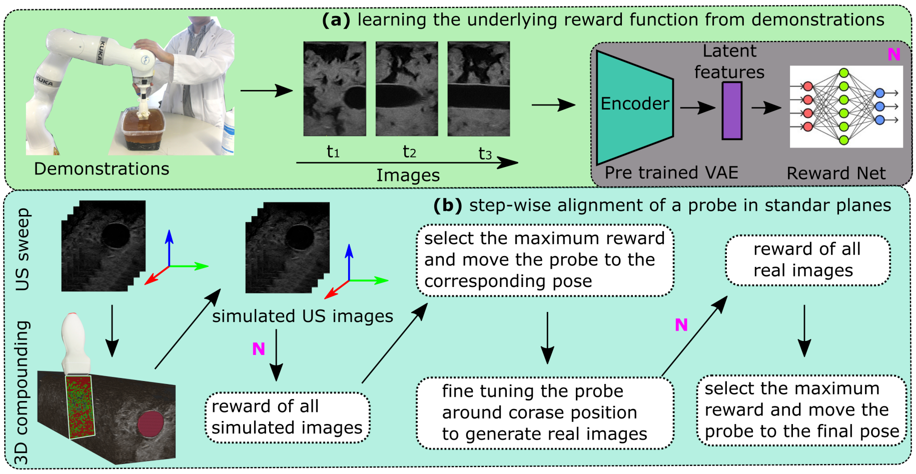


Fig. 6.2. Schematic overview of the framework.

In addition, to demonstrate the potential of the proposed approach, besides grid world analysis, the proposed method was validated on a gel vascular phantom and more challenging ex-vivo animal organs phantoms (chicken heart and lamb kidney). In contrast to simulation, such tasks are more realistic and more challenging in a physical environment because the resulting images are sensitive to practical factors, i.e., probe pose and contact force.

6.2.6 Contributions

In this work, we propose an advanced machine learning framework to explore standard scan planes defined by clinicians by demonstrations using robotic techniques. To understand the underlying intention and transfer the high-level physiological knowledge to the robotic system, a reward function is learned from image demonstrations to assess individual US images without any requirement of manual annotation. Such a system could enable consistent and repeatable diagnosis between different sonographers, particularly for inexperienced operators. To achieve this, a few demonstrations, guiding the probe from random poses (position and orientation) to the desired scanning plane, were recorded. To avoid biased results, a probabilistic metric was developed to assess the demonstrations and filter out abnormal ones using a joint Gaussian distribution. Since sonographers need to actively adjust probe pose to standard planes, these demonstrations are frequently and inherently sub-optimal. Inspired by [157], a self-supervised method based on the spatial cues of US frames in the demonstrations was developed, in which comparisons between two images are globally generated from all consistent demonstrations. Then the rewards for individual images can be inferred by a neural network trained using the pairwise image comparisons. In contrast to [157], and considering the time efficiency and variation of patient-specific situations, this work proposes a navigation process comprising coarse and fine adjustment steps instead of using RL or informative path planning approaches based on the computed reward function. To validate the performance of the proposed approach, experiments have been carried out in a simulated grid world and two typical cases on a gel vascular phantom, alongside more challenging ex-vivo animal organ phantoms (lamb kidney and chicken heart).

6.3 Learning from Few Demonstrations

This section describes the probabilistic temporal ranking approach (PTR) and the proposed GPSR method to learn the latent reward function for individual B-mode images from expert's demonstrations. In addition, a confidence-based approach was introduced to assess the quality of the given demonstrations to avoid biased end results.

6.3.1 Probabilistic Temporal Ranking (PTR)

PTR is a reward model that learns from pairwise comparisons sampled temporally from demonstrations [157]. This method models the rewards with a Gaussian process using a Matern32 kernel function. To train the model, a differentiable neural approximation was employed in [157]. More precisely, the Gaussian process was replaced with a single layer FCN with similar performance. The generative process for a pairwise comparison outcome ($g \in \{0,1\}$), between the latent rewards for the observations (images) r_{t_1} and r_{t_2} at time step t_1 and t_2, is modeled as a Bernoulli trial as in Eq. (6.1).

$$g \sim \text{Ber}(\text{Sig}(r_{t_2} - r_{t_1})) \tag{6.1}$$

where Ber represent Bernoulli distribution and Sig() represents the sigmoid function.

Regarding US scanning, PTR assumed that the images obtained at the later stage of a demonstration are more important than those obtained earlier. This assumption is reasonable for the applications searching for standard planes because sonographers usually stop and maintain the probe when the target planes have been achieved. According to this assumption, ground truth training labels $g = 0$ are generated if $t_1 > t_2$, and $g = 1$ if $t_1 < t_2$. The Bernoulli trial introduces slack that enables PTR to predict the reward from a sub-optimal demonstration, where the assumption above may not hold. This means the reward for the demonstration can be non-monotonic because there is a greater chance to get a comparison result accidentally $g = 0$ or 1 when two rewards are similar.

To robustly predict the rewards for unseen observations, a VAE [159] was used to extract the representative image features in latent space $Z_t \sim (\mu(X_t), \delta^2(X_t))$. X_t is the observation at timestamp t, μ and δ are the mean and diagonal covariance of the observations. The extracted latent feature Z_t was fed to an FCN to estimate the corresponding reward. To train PTR in an end-to-end fashion, the loss for VAE and reward FCN are combined as Eq. (6.2).

$$\begin{aligned}\mathcal{L} = &-E_{z_{t_1} \sim q(z_{t_1}|x)}\left[log\ p(x|z_{t_1})\right] + KL(q(z_{t_1}|x)||p(z_{t_1}))\\ &- E_{z_{t_2} \sim q(z_{t_2}|x)}\left[log\ p(x|z_{t_2})\right] + KL(q(z_{t_2}|x)||p(z_{t_2}))\\ &- \frac{1}{N}\sum_{i=1}^{N}\left[g_i log(h(g_i)) + (1 - g_i) log(1 - h(g_i))\right]\end{aligned} \tag{6.2}$$

where $h(g_i)$ is the computed result using Eq. (6.1) for a pair of observations and g_i is the ground truth label for i-th observation combination, which is automatically determined based on time sequence t_1 and t_2 of the two observations. By minimising the loss, the parameters for the encoder $q(z|x)$, decoder $p(x|z)$ and the reward net $r_\psi(z)$ can be obtained.

6.3.2 Global Probabilistic Spatial Ranking (GPSR)

PTR only considered the relationship between the frames in an individual demonstration, while the global relationship between demonstrations has not been investigated. Considering the potential inconsistency between demonstrations, it is important to rate the demonstrations, to automatically identify and filter undesirable data from the training set to avoid contradictory objectives hindering the learning of the demonstrator's latent intention.

Data cleaning

Regarding the task of visualizing the desired US scan planes, a demonstration ending at a good viewpoint, where the standard planes could be displayed, is considered a good demonstration. Since the resulting B-mode imaging depends on the end-effector pose, the end-effector poses of the last frames in all demonstrations were employed as the surrogate of the corresponding US images. Then, the last end-effector poses $\mathbf{P}_p \in R^6$ were modeled using a joint Gaussian distribution with N_{dof} (the number of DOFs) variables, and the parameters of the distribution were determined using a maximum likelihood method.

$$\mathbf{P}_p \sim \mathcal{N}(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g) \tag{6.3}$$

where $\boldsymbol{\mu}_g = \mathbb{E}(\mathbf{P}_p)$ and $\boldsymbol{\Sigma}_g$ are the mean vector and covariance matrix of $\mathbf{P}_p$ from all demonstrations, respectively. $\Sigma_{i,j} = \mathbb{E}[(P_p^i - \mu_g^i)(P_p^j - \mu_g^j)]$, $1 \leq i, j \leq N_{dof}$. To balance the numerical difference in different DOFs, $\mathbf{P}_p$ has been normalized into $[0, 1]$ in each DoF.

Regarding the multivariate normal distribution, the confidence interval for a certain significance level p_{ci} results in a region, in which $\mathbf{P}_p$ satisfies the following condition.

$$(\mathbf{P}_p - \boldsymbol{\mu}_g)^T \boldsymbol{\Sigma}_g^{-1} (\mathbf{P}_p - \boldsymbol{\mu}_g) \leq \chi^2_{N_{dof}}(p_{ci}) \tag{6.4}$$

where $\chi^2_{N_{dof}}$ is the percent-point function for probability p_{ci} of the chi-squared distribution with N_{dof} degrees of freedom. If the last pose of a sampled demonstration is not inside the confidence region with a given significance level (i.e., 5%), this demonstration is considered different from others and should be removed from the data set.

Probabilistic Spatial Ranking

After removing the unsuitable demonstrations, the latent objectives for different demonstrations are generally consistent. For PTR, the demonstrations are considered individually. However, if the paired images could be generated from different demonstrations, much more training data (pairwise comparisons) can be generated from few demonstrations. Considering US demonstrations are usually inherently sub-optimal and the demonstrations initialized at different positions, it is difficult to determine which image (from different demonstrations) deserves a higher reward based on temporal information alone. Without global pairwise comparisons, the difference between the last observations of all demonstrations will limit the achievement of a more consistent and precise result. In the worst case, the incorrect comparison outputs will further decrease the reward inference ability of the PTR approach. To overcome this drawback and enable the ability to generate images pairs with correct comparison output from different demonstrations, this work introduces GPSR, which uses the probe's spatial cue for image comparison instead of the temporal information used by PTR.

To visualize the tissues of interest in 2D standard planes, a US probe is required to be controlled in four DOFs for most applications, e.g., the task of visualizing and horizontally centering the lamb kidney and chicken heart (Fig. 6.1). Two translational movements in the plane orthogonal to the probe centerline and two rotational movements along the probe long axis and the probe centerline, respectively. The remaining two DOFs: the translation along probe centerline and the rotation around the probe short axis only change the visualized part of the same plane. In addition, to obtain high-quality US images, the probe is required to make firm contact with objects' surface. To consider the object's safety, a compliant controller is often used in RUSS to avoid too large force in the probe centerline direction. Besides, for some anatomies, i.e., tubular tissues, the translational position along the centerline of tubular tissues is allowed to be flexible to display a longitudinal vascular view.

For the proposed GPSR approach, the spatial cue of each frame is used as the supervisory signal instead of the time sequence to generate image comparisons between different demonstrations. Unlike PTR, we assume that the images obtained at the position close to the ending pose should be assigned a higher reward than the images obtained at the poses far away from the last pose in the demonstration. Considering that the optimal poses for the standard planes may vary a lot for some applications, like scanning the longitude view of tubular tissues, the absolute pose difference $\mathbf{d}_i^k$ between the pose of a frame and the pose of the last frame from the same demonstration, instead of absolute probe poses themselves, are employed by GPSR.

$$\mathbf{d}_i^k = \left[|\mathbf{P}_i^k(1) - \mathbf{P}_e^k(1)|, ..., |\mathbf{P}_i^k(6) - \mathbf{P}_e^k(6)| \right] \tag{6.5}$$

where $\mathbf{P}_i^k \in R^6$ and $\mathbf{P}_e^k \in R^6$ are the poses of i-th and ending frame of k-th demonstration.

To generate global comparisons between frames from different demonstrations, the computed $\mathbf{d}_i^k$ of all demonstrations were used to obtain the global maximum pose difference vector in all six DOFs $\mathbf{d}_{max}^g = \left[\max\left(d_i^k(1)\right), ..., \max\left(d_i^k(6)\right) \right]$. To balance the numerical difference between translational movements and rational movement, $\mathbf{d}_i^k$ was further normalized to $[0, 1]$ in each DOF, respectively, as follows:

$$\mathfrak{D}_i^k = \left[\frac{\mathbf{d}_i^k(1)}{\mathbf{d}_{max}^g(1)}, ..., \frac{\mathbf{d}_i^k(6)}{\mathbf{d}_{max}^g(6)} \right] \tag{6.6}$$

The global generalized distance D_i^k was calculated as Eq. (6.7).

$$\begin{gathered} D_i^k = \sqrt{\sum_{j=1}^{6} \left[k_j \left(\mathfrak{D}_i^k(j) \right)^2 \right]} \\ \text{where } k_j = \begin{cases} \dfrac{\left(\mathbf{d}_{max}^g(j)\right)^2}{\sum_{m=1}^{3} \left(\mathbf{d}_{max}^g(m)\right)^2} k_t & j = 1, 2, 3 \\ \dfrac{\left(\mathbf{d}_{max}^g(j)\right)^2}{\sum_{m=4}^{6} \left(\mathbf{d}_{max}^g(m)\right)^2} k_r & j = 4, 5, 6 \end{cases} \end{gathered} \tag{6.7}$$

where k_t and k_r are the weights for translational and rotational movements, respectively, $k_t + k_r = 1$. k_j is the weight for the movement j-th DOF, which is automatically determined based on the demonstrations. The DOF with larger movement variation among the demonstrations will result a larger weight.

Then global comparison outcomes are determined by the global generalized distance D_i^k. For pairwise images randomly selected from demonstrations, the comparison outcome label $g = 1$ is assigned if $D_i^k(p_1) \geq D_i^k(p_2)$, and $g = 0$ if $D_i^k(p_1) < D_i^k(p_2)$. Similar to PTR, the latent features from the VAE encoder were used as the input for an FCN reward prediction network $r_\psi(Z)$. The detailed architecture has been shown in Fig. 6.3. To determine the weights of FCN, a binary cross entropy loss over the comparison outcome was used as reward loss function $\mathcal{L}_{re}$.

$$\mathcal{L}_{re} = -\frac{1}{N}\sum_{i=1}^{N}[g_i log(h(g_i)) + (1-g_i)log(1-h(g_i))] \tag{6.8}$$

where $h(g_i) = \text{Sig}(k_{re}(r_{p_2} - r_{p_1}))$, $r_{p_1} = r_\psi(Z_{p_1})$ and $r_{p_2} = r_\psi(Z_{p_2})$ are the predicted reward for the images obtained at p_1 and p_2, respectively, and k_{re} is the coefficient used to restrain false high reward. k_{re} was empirically set as 5 in this work.

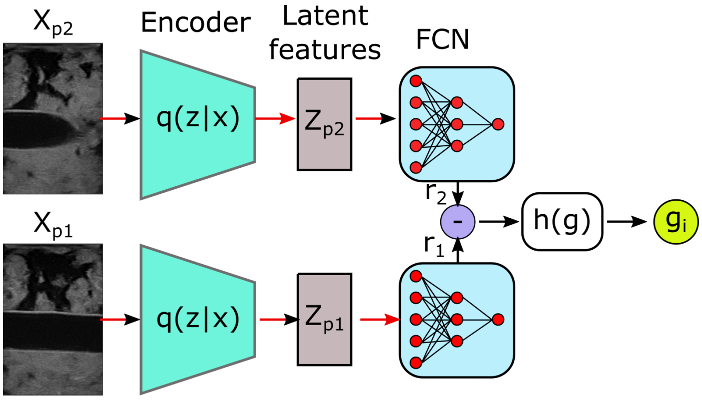


Fig. 6.3. The architecture of the GPSR reward net. The paired images are encoded by a pre-trained VAE. Then, the image features (Z_p) in latent space are fed to an FCN to predict the reward for the current images. Based on the difference between the predicted rewards, a comparison outcome can be generated using $h(g)$.

6.4 Autonomous Navigation to Standard Planes

6.4.1 Generation of Simulated US Images

Automatic Scanning and US Volume Reconstruction

Given the trained reward FCN $r_\psi(Z)$, it becomes possible to evolve a policy to navigate the US probe to target standard planes using RL methods or Gaussian processes [157]. However, to find an optimal policy, many interactions with subjects are required to balance both exploration and exploitation. Due to patient safety concerns, these approaches are not suited for US applications. In addition, time efficiency is another practical issue because it is not practical to ask patients to stay stationary for a long time during an examination.

To address such practical limitations in US examination, a new framework is proposed to quickly and safely navigate the US probe to the target scan planes. Instead of real-time interaction with subjects, a single sweep over the tissue of interest was carried out. To achieve this, three points ($\mathbf{P}_i^m$) are manually selected on the object surface in a clockwise or anti-clockwise direction. The anatomy of interest should be covered by the parallelogram defined using $\mathbf{P}_i^m$. To obtain the 3D reconstruction of the area of interest, a scan path planning method similar to [57] was employed here. Since the probe width w_p is limited, a multi-line trajectory is used to completely cover the area of interest as Fig. 6.4. To minimize the total length of the trajectory, the probe is mainly moved in the direction with a larger length, i.e., $\overrightarrow{p_2^m p_1^m}$ when

$|p_2^m p_1^m| \geq |p_2^m p_3^m|$. The key points $\mathbf{P}_k$ required to define the multi-line trajectory are calculated as follows:

$$\mathbf{P}_k = \begin{cases} \mathbf{P}_2^m + \frac{1}{2}\overrightarrow{p_2^m p_i^m} & w_p \geq \min\{|p_2^m p_1^m|, |p_2^m p_3^m|\} \\ \mathbf{P}_2^m + \left[\left(\frac{1}{2} + j\right) w_p - j\epsilon_0\right] \frac{\overrightarrow{p_2^m p_i^m}}{|\overrightarrow{p_2^m p_i^m}|} & \text{Others} \end{cases} \tag{6.9}$$

where $\overrightarrow{p_2^m p_i^m}$ is the axis with a shorter length, $i = \{1, 3\}$, $j = 1, 2, ..., \lceil \frac{\min\{|p_2^m p_1^m|, |p_2^m p_3^m|\}}{w_p} \rceil$, ϵ_0 is a small coefficient used to guarantee the overlap between the sweep along two neighbouring lines.

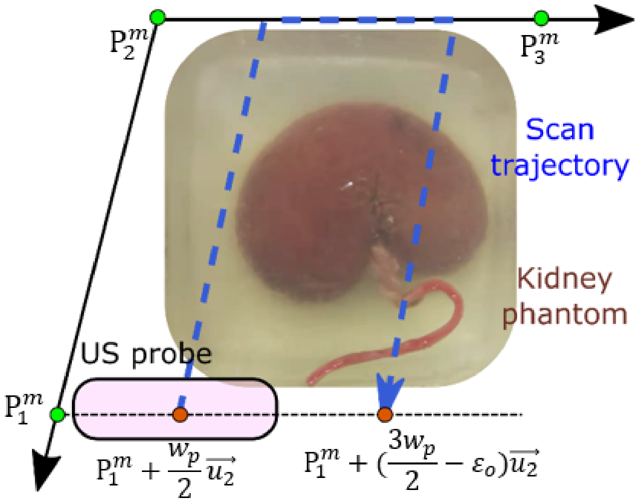


Fig. 6.4. The transformation between coordinate systems.

After accomplishing the sweep along the planned trajectory, the tracked B-mode images can be stacked in spatial space based on the known robotic kinematic model. In order to obtain the 3D volume, a linear interpolation was employed, using PLUS [160].

Simulated US Images from Real Volume

In order to obtain 2D images from the compounding result, a rectangle ($w_p \times d_p$) was employed, where d_p is the depth of the US setting. Then the volume was projected into $X - Y$ plane of the robotic base coordinate system {b} (Fig. 6.5), in which the probe pose could be directly used for the controller. Then the probe pose in the 2D view could be defined as (${}^bx$, ${}^by$, and b_z^R). In case the desired plane is only able to be seen when the probe is tilted, additional rotation around probe long axis (${}^pX$) ${}^pR_x$ was defined. Since the effect of the rotation around the probe short axis (${}^pY$) could be partly replaced with the movement in ${}^pY$ direction, this was kept constant.

Then, a large number of simulated images could be obtained by giving different (${}^bx$, ${}^by$, ${}^bR_z$, ${}^pR_x$). The policy employed to generate simulated images here is: 1) ${}^bx = [x_{min}, x_s, x_{max}]$, 2) ${}^by = [y_{min}, y_s, y_{max}]$, 3) $b_z^R = [0, R_z^s, 180°]$, and 4) $b_x^R = [-30°, R_x^s, 30°]$. The left, middle and right values represent lower boundary, the increment of each step, and the upper boundary, respectively. In this work, $x_s = y_s = 3\ mm$ and $R_z^s = R_x^s = 5°$.

To filter out the simulated images without correct context, the volume mask M_v (2D) was generated based on the 3D compounding result, which is the projected result of 3D volume in $X - Y$ plane of frame {b} (Fig. 6.5). Regarding M_v, the white area ($I = 1$) represents the projected volume with context inside, while the black area ($I = 0$) is the area padded during 3D compounding. Then, another mask for probe M_p could be featured as a line defined by (${}^bx$, ${}^by$, and b_z^R). Considering the width of the probe is w_p, M_p can be generated as a line with high intensity ($I = 1$) while the intensity of other area are zero. The line is represented by $(x, y) = \{(x, y) | \frac{y - {}^by}{x - {}^bx} = \tan(b_z^R)$ && $x \in [{}^bx - \frac{w_p}{2} \cos b_z^R, \ {}^bx + \frac{w_p}{2} \cos\{b_z^R\}\}]\}$. Then, an element-wise multiplication was applied between two masks ($M_v \odot M_p$). If the probe is not fully located inside of the volume mask, the corresponding simulated images will be discarded.

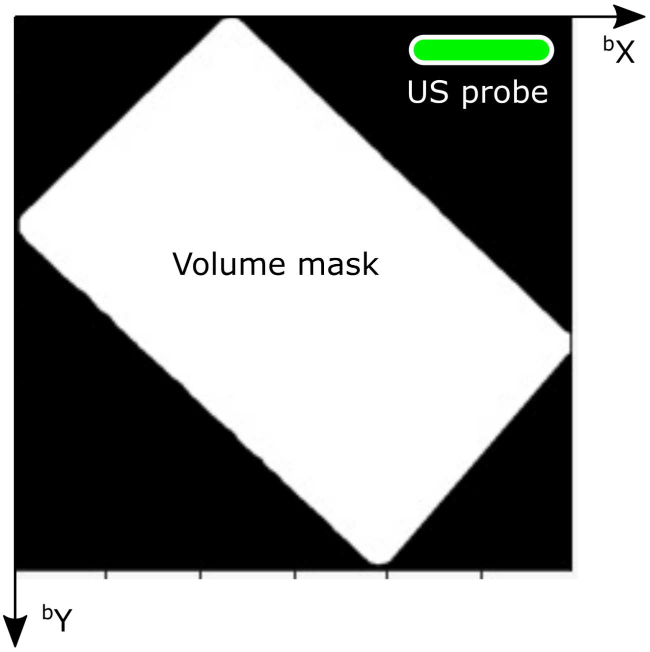


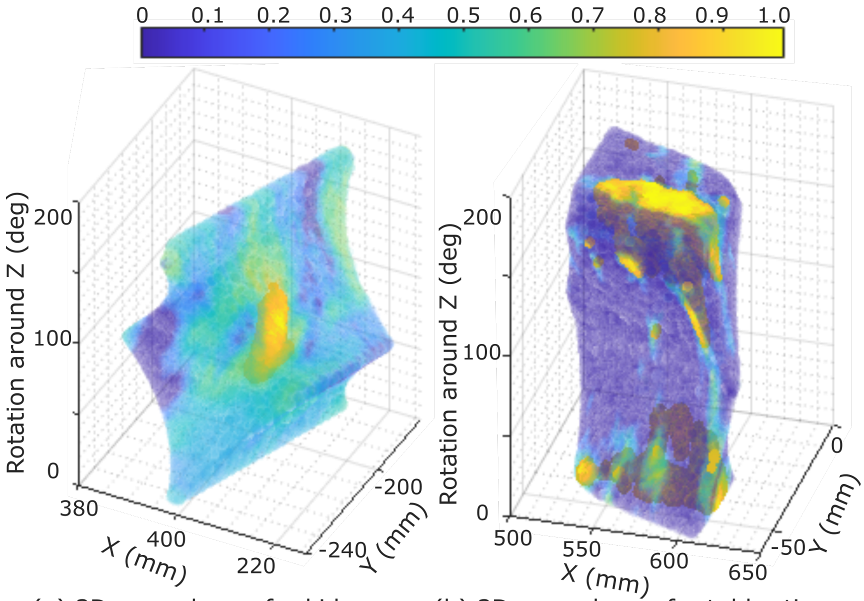
Fig. 6.5. The transformation between coordinate systems.

6.4.2 Alignment of US Probe

Coarse Positioning of US Probe

After applying the well-trained reward network $r_\psi(Z)$ for the simulated US images, the reward volume with respect to the variables can be generated. To intuitively visualize the reward distribution, two representative reward volumes with respect to the probe positions in X and Y direction and the rotation around Z direction (${}^bx$, ${}^by$, and ${}^bR_z$) for tubular phantom and ex-vivo kidney phantom, respectively, are shown in Fig. 6.6. The dark blue area represents low reward values, while the yellow area represents high reward. It can be seen from the figure that the high reward (> 0.85) is mainly distributed around the area ($mean \pm STD$) $[410 \pm 0.8\ mm, -219 \pm 1.8\ mm, 106.5 \pm 3.1°]$ for kidney phantom and $[579 \pm 17.1\ mm, -38 \pm 3.3\ mm, 170 \pm 6.4°]$ for the tubular phantom. For most applications, i.e., the lamb kidney and chicken heart, the highest rewards are concentrated to a point, while for tubular tissue these are mainly distributed along a line. This is because of the different geometric features of the anatomies involved; the longitudinal view of the blood vessel allows flexibility in the direction of the vessel centerline.

Considering the efficiency of the approach, the generated simulated images were fed into the reward FCN. Then, a 3D moving window was used to smooth the 3D volume of the rewards. Based on the resulting rewards, the mimicked image with maximum reward was extracted and we directly move the probe to the corresponding pose.



(a) 3D reward map for kidney (b) 3D reward map for tublar tissue

Fig. 6.6. The transformation between coordinate systems.

Fine Adjustment of US Probe

Like human operators performing US probe examinations, a fine-tuning process was performed after coarse positioning of US probe. Considering the size of objects, a searching area ($^bx : \pm 10\,mm$, $^by : \pm 10\,mm$, $^bR_z : \pm 10°$, and $^pR_x : \pm 10°$) was determined around the target pose using the simulated images. The step size for the involved translational movement was $2\,mm$ and the one for the rotational movement was $5°$. Then, the searching procedure was automatically performed in the specified area. During the fine-tuning procedure, the reward for each image was computed in real-time. The US image with the maximum reward was finally considered the best estimation of the target standard plane. In addition, all searching parameters used in the fine-tuning procedure can be changed according to the requirements of different tasks.

6.5 Results

6.5.1 Hardware Setup

The proposed RUSS consists of a robotic arm (KUKA LBR iiwa 7 R800, KUKA Roboter GmbH, Augsburg, Germany) and a US machine (Cephasonics, California, USA). A linear probe

(CPLA12875, Cephasonics, California, USA) was rigidly attached to the robotic flange using a custom-designed holder. The B-mode images were accessed using a USB interface provided by the manufacturer in $50\ fps$. The robotic arm was controlled using a Robot Operating System interface [48]. The real-time US imaging and robotic pose were synchronized and further used for 3D compounding in a software platform (ImFusion GmbH, Munich, Germany). The US setting was mainly determined by the default setting file for vascular tissues provided by the manufacturer: brightness: $66\ dB$, dynamic range: $88\ dB$. The depth, focus and frequency were $5.0\ cm$, $3.1\ cm$ and $7.6\ MHz$ for vascular phantom, while these parameters were changed to $5.5\ cm$, $5.0\ cm$ and $5.7\ MHz$ for the ex-vivo phantoms.

To validate the performance of the proposed approach for various organs, a blood vessel phantom and two ex-vivo organs phantoms (lamb kidney and chicken heart) were used in this work. The blood vessel phantom was primarily made using gelatin powder, while the ex-vivo organs phantoms were made using candle wax. Candle wax has a good sealing ability, which allows for longer preservation of fresh animal organs. To construct the vascular phantom: gelatin power ($175\ g/L$) was dissolved into water, and the mixed solution was heated to 80 degree. To mimic human tissue artifacts, paper pulp ($3-5\ g/L$) was randomly mixed into the solution. After solidification, a round tube was used to create two holes at different depths of the phantom for mimicking vascular structures. For the ex-vivo organs phantom, hot candle wax liquid was used to cover the organs in a small box. After the liquid was fully solidified, it was taken out and placed in another box. Then, candle wax liquid was poured into the box to submerge the upper surface of the phantom. Like paper pulp used for vascular phantom, ginger power ($10\ g/L$) was mixed with the candle wax for ex-vivo organ phantoms.

6.5.2 Performance of GPSR on 2D Grid World

To theoretically validate whether the proposed GPSR method can properly learn the reward model interpreting the underlying human intention, a grid world environment (20×20) was built for qualitative and quantitative analysis. In the grid world, the aim is to move an agent initialized from a random position towards the target position. To simulate the two most typical tasks in clinical routine, a point and a line object were employed, respectively, to mimic the tasks require strict and non-strict probe pose to repeatedly show object standard planes.

In order to generate sub-optimal trajectory demonstrations, two q-learning policies were trained separately for the point and line objects in the grid world, in which the reward of the target position was set to one while the rewards of other positions were zero. The maximum episodes and steps for each episode were set to 50 and 100, respectively. Based on the sub-optimal trained model, an agent will move towards the target position from a random position. For point object, five demonstrations ending at the target position were generated to mimic standard planes in US applications. In addition, for the line object (1×5), ten demonstrations ending at the target line were generated. Since the model was only trained with limited epochs, the generated demonstration was sub-optimal; namely, the trajectory cannot directly move towards the target position in the minimum number of steps.

Based on the given sub-optimal demonstrations, three reward models were trained using maximum entropy IRL (ME-IRL) [155], PTR [157] and the proposed GPSR, separately. The

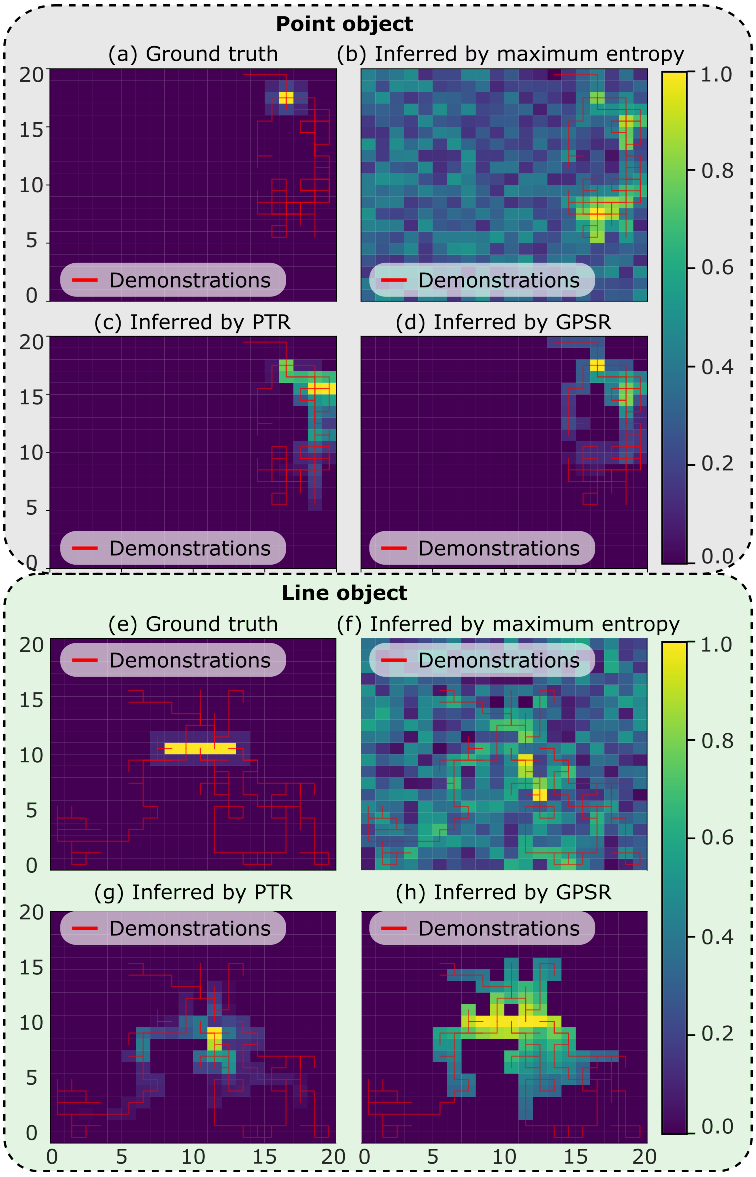


Fig. 6.7. Reward maps obtained from sub-optimal demonstrations. Red lines represent the demonstrations, and the color in each grid shows the reward of the corresponding position.

ground truth and the reward maps inferred by the three models are shown in Fig. 6.7. It can be seen from the figure that the reward map inferred by the proposed GPSR is closest to the ground truth, in which the highest reward clearly shows in the right position in the grid world. The result achieved using the maximum entropy [155] approach failed to correct the recover the correct reward value from the sub-optimal demonstrations, particularly for the line object. This is mainly because of the inherited limitation of the maximum entropy approach, which tends to assign larger reward values to states observed more often in the demonstration. However, demonstrations of US examination are sub-optimal. The positions observed more often in the demonstrations may not be the desired positions. Regarding PTR approach [157], the inferred reward map is much better than the one obtained from maximum entropy while still worse than the one achieved from the proposed GPSR approach. It can be seen from Fig. 6.7 (c) that the position with the maximal reward is close to the desired position. However, since the pairwise training data was generated from individual demonstrations, a biased result occurs, especially when the lengths of demonstration are significantly different from each other. For the line object, PTR fails to realize that the object is a line. Only one grid point of the line is successfully assigned with a large reward (see Fig. 6.7 (g)).

Considering the intrinsic properties of US examination, the target standard planes correspond to a unique or a set of probe poses with a unified character. Thus, the proposed GPSR approach directly generates the paired training data from all demonstrations based on the probe spatial cues rather than temporal cues. Due to the global consideration of all demonstrations, more training data could be generated from the same demonstration using the GPSR approach. The performance of inferred reward in grid world using different approaches also demonstrates the superiority of GPSR over the PTR and maximum entropy approaches, particularly for unstructured objects like a line object (Fig. 6.7).

To further quantitatively compare the performance of different approaches, four q-learning models were trained, separately, according to the three inferred reward maps and the ground truth. The number of training episodes was set to 500, and the maximum number of steps for each episode was 100. Then, 20 points were randomly initialized in the grid world, and the four trained models were used to guide the agents to move towards the target position separately. Considering time efficiency, the trials were only considered successful when the agent can stop at the desired position within 100 steps. The whole process was repeated 20 times with random target positions. The final success rate over 400 trials (20×20) for different approaches is summarized in TABLE 6.1.

It can be seen from TABLE 6.1 that the ME-IRL results in the worst case in our setup, where the success rate is only 9.3% and 3.2% for the point and line objects, respectively. In contrast, the result obtained using the ground truth is best (99.3% and 100%). In addition, the success rate achieved using the PTR and the proposed GPSR are 47.0% and 72.0% for the point object, respectively. Although GPSR results are still worse than the ground truth, the performance already significantly improved from the state-of-the-art approaches (62.7% and 25%, respectively). In addition, regarding the line object, the success rate for PTR further decreases to 33.2% while the GPSR increase the number to 84.5%. This is mainly because PTR only considers the time sequence of individual demonstrations, while the proposed GPSR globally generates pair-wise comparisons between all available demonstrations.

Tab. 6.1. Performance of Learning from Demonstrations Methods

Methods	**Success rate**	
	Point object	Line object
Ground truth	99.3%	100%
ME-IRL	9.3%	3.2%
PTR	47.0%	33.2%
GPSR	72.0%	84.5%

6.5.3 Performance of Rewards Inference Model

Training Details

For each phantom, ten demonstrations were given by experts individually. The lengths of demonstrations vary from 871 to 1797 frames depending on the initial positions and expert behavior. Differing from the end-to-end mode used in [157], here the VAE was trained individually to guarantee the consistent performance for the simulated images, which are slightly different from acquired images due to the interpolation during 3D compounding. To train the VAE, the data set of each phantom consisted of 6000 US images randomly extracted from the demonstrations and 2000 simulated images from 3D volumes. The ratio between the training and validation data sets was $9:1$. The size of the input images is 256×256 and the size of the latent feature was set to 64. The batch was set as 64 and the total training epoch was set to 100. The network parameters were optimized using Adam with a constant learning rate 1×10^{-3} [112].

Regarding the training process of the proposed GPSR approach, only five demonstrations were randomly selected. Four of them were used as training data set, while the remaining one was used for validation. To balance the weights of different demonstrations, all the demonstrations were down-sampled to 500 frames. Therefore, the size of the pair-wise training data set is $C(2000, 2) = 1,999,000$, while the size of the validation data set was $C(500, 2) = 124,750$. Then, the reward network was trained by minimizing the cross entropy loss described in Eq. (6.8). The training details are as follows: batch size: 128, epoch: 5, leaning rate: 1×10^{-5}. In addition, all the training process was run on a single GPU (Nvidia GeForce GTX 1080).

The training loss and validation loss of the reward interference network for vascular phantom are used as representative results in Fig. 6.8. Both training loss and validation loss can effectively converge to stable results after $8,000$ iterations.

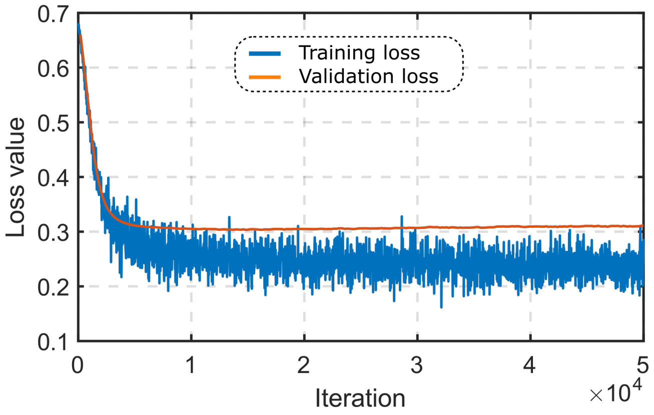


Fig. 6.8. Training and validation loss of the reward model for vascular phantom.

Inferred Reward

To intuitively validate whether the learned reward model can learn the underlying skills from the expert demonstrations, the learned reward model was used to estimate the reward of each frame among three unseen demonstrations for the vascular phantom, chicken heart, and lamb kidney phantoms, respectively. According to the description of GPSR, the desired reward is negatively correlated to the global generalized distance D_i^k. For better visualization, the inverse global generalized distance $(1 - D_i^k)$ was drawn together with the estimated reward purely based on the US images in Fig. 6.9.

In general, the inferred rewards for different anatomies have a consistent relationship with the corresponding inverse global distance. Compared with beginning frames, a much higher reward is achieved by the frames at the end of the demonstrations, which is consistent with the underlying phenomena for the discovery task that the demonstration will be ended when the target plane has arrived. Regarding the data from the experiments on the chicken heart and kidney phantom, it is notable that there are few frames, noted as B2 and C2 in Fig. 6.9, with high rewards in the middle of the demonstration. Compared with the images acquired in the last frame of the demonstration, the corresponding images obtained at B2 and C2 are very similar to the images acquired at B3 and C3. This can be explained by the underlying manner that how human operators do a discovery task: they usually can quickly guide the probe to a coarse pose around the desired one. Then, a time-consuming fine-tuning procedure is performed to identify the best one, although the globally optimal one may already be obtained in an earlier phase. The time for fine-tuning procedure increases when the image is more challenging; for example, the time consumed by the fine-tuning procedure over the whole process is around 20% for the mimic vascular phantom, while the time increases to around 50% for the more challenging ex-vivo animal tissues.

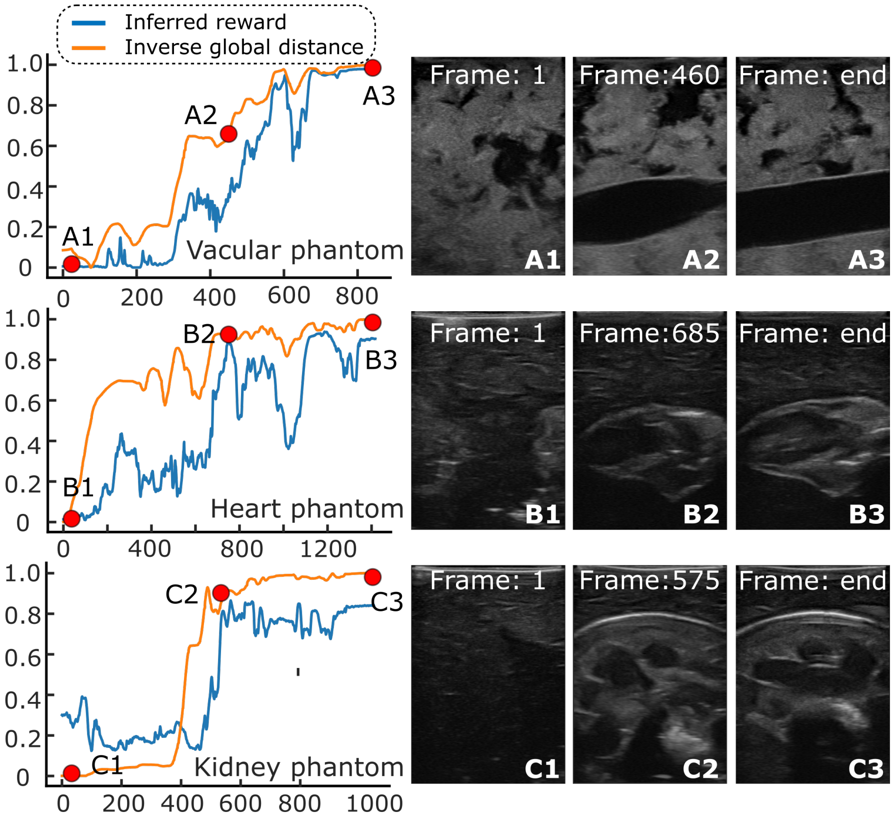


Fig. 6.9. Reward maps obtained from sub-optimal demonstrations. Red lines represent the demonstrations and the color in each grid shows the reward of the corresponding position.

6.5.4 Performance of Standard Scan Planes Alignment

In order to validate the proposed pipeline in realistic cases, the longitudinal view of blood vessels and ex-vivo animal phantoms was used as the target plane. To further validate whether the proposed method can deal with an unseen case, two blood vessels located at different depths (vessel 1: $26\ mm$ versus vessel 2: $37\ mm$) of the custom-designed gel phantom were used. In addition, to compare the performance between the proposed approach and a human operator, the variations on final position and orientation for different experiments were recorded. Finally, considering that objects could be moved after training, the phantoms were rotated into different angles ($30,\ 60,$ and $90°$) to validate whether the method was still able to place the probe in the target planes correctly.

Comparison with Human Operators

To quantitatively compare the performance between human operators and the proposed GPSR approach, ten experiments were performed by a well-trained human operator and the proposed GPSR approach, respectively, to position the probe along the longitudinal plane of the vascular phantom, namely, vessel 1. The mimicked blood vessels are straight holes inside the phantom, allowing positional slack in the direction of the vessel centerline. Thus, the distance (e_d) between the vessel centerline and the centroid of the cross-sectional vessel images in 3D space was defined instead of absolute position. The centroid of the vessel in US images was calculated using the same steps as [7]: (1) using a U-net to segment the vessel from US images, (2) computing the centroid based on the binary map using OpenCV, and (3) applying the spatial calibration result to calculate the 3D position of pixel-wise centroid in 2D images. The ground truth of the vessel centerline was computed using the same method by moving the probe along the vessel centerline. Besides metric e_d, the absolute rotation error (e_r) between the probe long axis (${}^{p}X$) and the vessel centerline was further defined to assess the probe orientation.

The results of e_d and e_r over the 20 trials (ten for each) performed by human operators and the proposed GPSR approach on vessel 1 are summarised in Fig. 6.10. Regarding e_d, GPSR results are more concentrated and the average e_d ($6.6 \pm 0.1\ mm$) is less than the one obtained from human operators ($9.1 \pm 1.7\ mm$). Regarding e_r, the proposed GPSR achieves comparable results ($1.5 \pm 1.6°$) to the human operator ($1.3 \pm 1.0°$). Based on a significant test (t-test), the probability between e_r obtained from human operators and the GPSR is $0.72 > 0.05$, which means there is no significant difference between these data.

To further validate the generalization ability, an unseen blood vessel, namely vessel 2, was employed to represent a patient-specified object. To compare the manual performance for different vessels, ten independent trials were carried out by the same human operators. Without further updating the parameters in the trained reward network using demonstrations for vessel 1, the GPSR also ran ten times on vessel 2. The results have been shown in Fig. 6.10. In terms of e_r, the human operators have very close results to each other ($1.5 \pm 1.6°$ and $1.7 \pm 1.8°$) while e_d obtained for vessel 2 is smaller and more concentrated than the one of vessel 1 ($6.2 \pm 0.9\ mm$ verse $9.1 \pm 1.7\ mm$). This is mainly because human operators are limited to perception accuracy, and different biases may occur for different objects. Regarding the performance of the proposed GPSR approach on an unseen vessel 2, e_r ($3.8 \pm 0.9°$) is slightly larger than that obtained by humans while the smallest e_p ($4.1 \pm 0.2\ mm$) is achieved among all groups. This means that the GPSR has the potential to address the patient-specific case in a real scenario. In addition, it is also noteworthy that the GPSR can generate better results than individual demonstrations, i.e., e_d. This is because the demonstrations were filtered to remove poor quality examples, and since demonstrations are considered globally rather than individually as in [157].

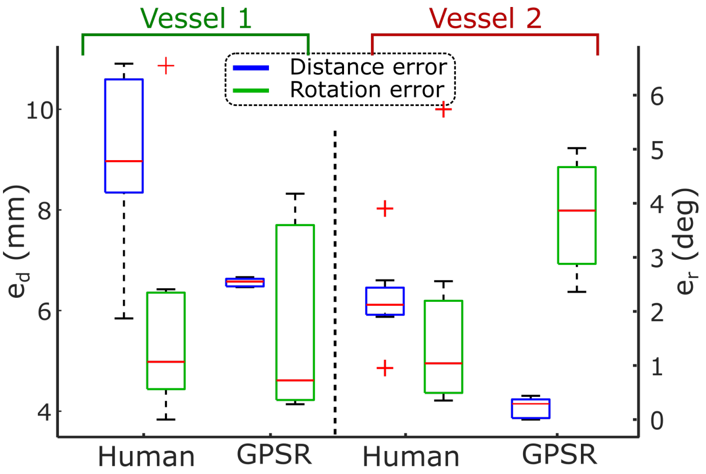
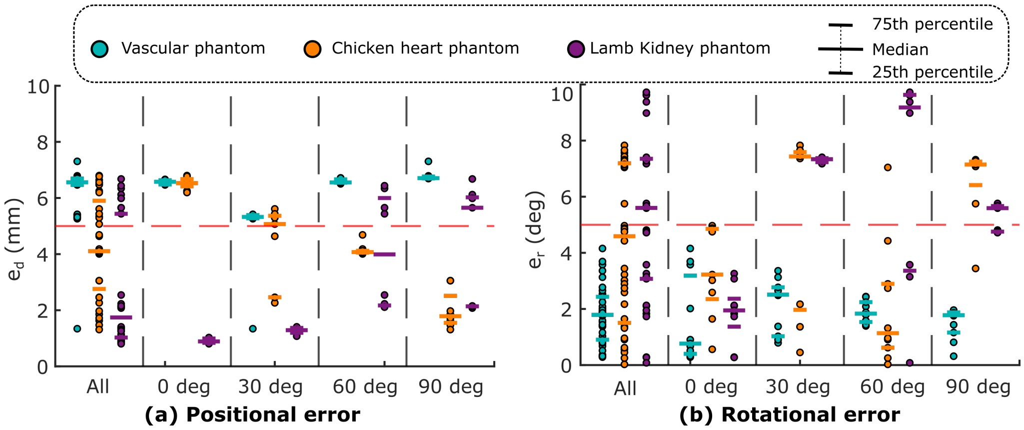


Fig. 6.10. The performance of human operators and the proposed GPSR approach on two different vessels.



(a) Positional error

(b) Rotational error

Fig. 6.11. Reward maps obtained from sub-optimal demonstrations. Red lines represent the demonstrations and the color in each grid shows the reward of the corresponding position.

Quantitatively Assessment of Standard Planes Alignment

This section demonstrates the quantitative results of the proposed GPSR approach on different objects (tubular tissues, chicken heart, and lamb kidney phantoms). To validate whether the GPSR has the potential to be used in a real scenario, the phantoms are placed in different positions and orientations during the experiments. This is because patients' positions for different examinations are different in clinical practice. For each phantom, four groups of experiments with different angles ($0, 30, 60,$ and $90°$) rotated around a given point on the flat table were carried out. Each group consisted of ten independent experiments.

The performance of the final alignment was assessed using positional (e_d) and rotational (e_r) errors as in the last subsection. However, differing from the e_d defined for vascular phantom, the absolute positional error between the final position achieved using the GPSR approach and the ground truth was used for kidney and chicken heart phantom because the standard planes are only able to be obtained by a unique pose. The ground truth of each pose was manually determined by the mean of ten demonstrations. The final performance is summarized in Fig. 6.11.

In general, e_d for vascular, chicken heart and lamb kidney are $6.2 \pm 1.0mm$, $4.3 \pm 1.8\ mm$ and $2.7 \pm 2.1\ mm$, respectively. Besides, e_r for the three different phantom are $1.8 \pm 1.0^{\circ}$, $4.3 \pm 2.7^{\circ}$, and $5.2 \pm 2.9^{\circ}$, respectively. Compared with the results on ex-vivo phantoms, the average e_d of the vascular phantom is slightly larger than others, while the average e_r of the vascular phantom is smaller than its peers. These results are mainly caused by the different geometries of objects. The tubular structures allow positional slacks along the centerline while the probe orientation is necessary to be accurately repeated to achieve target standard planes.

In addition, it can be seen from Fig. 6.11 that the trained GPSR model can be applied to the phantoms located in different positions and orientations without further training. In the worst case, the maximum e_d and e_r for all phantoms are $7.3\ m$ and 9.7° while the best results can achieve $1.3\ mm$ and 0.3° when the phantoms were rotated 90°. This means that the proposed approach can automatically identify the standard planes for challenging ex-vivo animal organs even when the object's position and orientation change. This ability enables the possibility of addressing the practical factors, i.e., patients' position variation, for different trials in a real scenario.

Besides the spatial accuracy of the final pose, the time required for the main components of the proposed GPSR approach for identifying the target standard planes are also summarized as follows: the simulation process costs $244\ s$ for generating $4,643$ images ($52\ ms$ for each image), and the reward network costs $11 \pm 4\ ms$ for computing the reward for each image. The time efficiency of the proposed approach makes it more practical to be used in clinical practice.

6.6 Conclusion

In this work, we present an advanced machine learning framework to automatically discover standard planes based on a limited number of expert demonstrations. To address the challenges of inevitable sub-optimal US examination demonstrations, the self-supervised GPSR approach was used to infer the reward for individual US images by understanding the underlying clinicians intention. The ability to learn from demonstrations can contribute to the worldwide deployment of RUSS because it allows sonographers to intuitively transfer their high-level clinical knowledge to RUSS without any requirement for robotic programming skills. It is our hope that this would lead to more consistent and repeatable diagnosis being achievable among different sonographers, particularly for inexperienced operators.

To validate the proposed approach, experiments were performed in a grid world and physical gel phantoms, respectively. In the grid world, the experimental results clearly demonstrate that the proposed GPSR is superior to the state-of-the-art approaches (maximum entropy [155] and PTR [157]) to generate a closer reward map to the ground truth. In addition, the experimental results on two typical clinical tasks on a tabular phantom allowing slack along the centerline and challenging ex-vivo animal organ phantoms (chicken heart and lamb kidney) strictly requiring location accuracy are $6.2 \pm 1.0mm$, $4.3 \pm 1.8\ mm$ and $2.7 \pm 2.1\ mm$ in terms of e_d and $1.8 \pm 1.0^\circ$, $4.3 \pm 2.7^\circ$, and $5.2 \pm 2.9^\circ$ in terms of e_r, respectively. In the future, the proposed intelligent robotic sonographer for discovering and locating standard scan planes can further enable the development of a fully automatic intervention system (i.e., vessels [138]) and extensive US examination programs for the early diagnosis and monitoring of internal lesions or tumor.

Part III

Conclusions and Outlook

7 Summary and Findings

In this thesis, several advanced robotic US imaging techniques were developed towards the aim of intelligent robotic sonography. Each technique can address the challenges encountered in clinical routine or in further improving the level of autonomy. These novel characteristics of a robotic US imaging system were developed based on interdisciplinary knowledge of robotic control, computer vision, and also clinical background. In this Chapter, we present the overall summary of each work.

7.1 Robotized US Probe Orientation Optimization

In Chapter 2, we presented two distinct approaches to automatically orienting US probes in the normal direction of unknown constraint surfaces, improving the sound propagation and enabling RUSS to reach predefined orientations relative to the surface normal at the contact point. Besides, the ability to robustly orient probe pose on patients is one of the dominating factors for accurate and reproducible US examinations. Since more acoustic signals can be reflected back to the transducer, rather than scatter away, the anatomy boundary can be improved in B-mode images, particularly for bone structure. Thereby, such a study can further enable accurate transferring of the planned operation trajectory from a preoperative image to real scenarios in orthopedic surgeries.

Both methods achieved accurate estimations of the normal direction by tactile and imaging information, compared with the camera-based approach ($7.3 \pm 3.1^\circ$) [56]. Due to use of an accurate F/T sensor, the performance of the mechanical model based approach is slightly better than the image and force based approach (in-plane: $2.7 \pm 1.9^\circ$ vs $3.1 \pm 1.0^\circ$; out-of-plane: $1.7 \pm 0.9^\circ$ vs $2.9 \pm 0.2^\circ$). Yet, the cost of image and force based approach is higher than the other approach as well. In addition, the mechanical model based approach is applicable both for the most common two types of probe (convex and linear), while the image and force based approach is more suitable for convex probe due to the use of B-mode images. The experiments were systematically performed on both gel phantoms and in-vivo tissues (forearm, upper arm, and back).

7.2 Motion-Aware Robotic US Imaging

In Chapter 3, we described the novel vision-based motion-aware approaches aiming to achieve accurate and consistent 3D images of target anatomies when an object motion is in presence during US scanning. In practice, human sonographers react to patient movements by repositioning the probe or even restarting the acquisition. Furthermore, several attempts for adjustment of objects are often necessary to clearly and completely image the anatomy of interest, e.g., adjusting limbs to acquire images of their entire artery tree. Thus, to improve the clinical acceptance of robotic US imaging systems, the motion-aware ability is important.

To monitor the tissue motion and properly compute the corresponding compensation,the passive marker based approach is first presented for feasibility validation. In this method, a depth camera is used to extract the manually planned sweep trajectory. Subsequently, to monitor the movement and further compensate for this motion to accurately follow the trajectory, the positions of firmly attached passive markers on the surface of a gel phantom were tracked in real-time. The compensation is computed using the ICP algorithm based on the markers positions before and after the movement.

To further explore the effective pipeline in real scenarios, the marker-less approach is developed based on surface registration. The US scan trajectory is extracted from a general CT atlas, in which the target object has been segmented. The motion monitoring system is based on the real-time segmented object masks obtained from RGB images. Once the subject moves, the robotic US will stop and automatically update its trajectory by registering the surface point clouds extracted from a depth camera before and after the movement using the ICP algorithm. Afterward, to ensure contact condition after repositioning of US probe, a confidence-based orientation optimization is employed to fill in the potential gap between probe and contact surface. Finally, the whole system is fully validated on a human-like arm phantom with an uneven surface.

The results demonstrate that these two approaches both work boldly in their individual environments. Regarding the passive marker based approach, the compensation accuracies are $3.1 \pm 1.0\ mm$ and $2.9 \pm 0.7\ mm$ of translation movement and rotation movement, individually, for the translation of up to $200\ mm$ and rotations of up to 40°. Besides, the markerless approach was validated on a human-like arm phantom with an uneven surface. The results demonstrate that the system can automatically resume a sweep when the subject moves during scanning. Compared with the marker based approach, a markerless approach is more convenient for operators without the requirement for carefully configuring the markers for individual patients. Such improvement we believe can bring them closer to clinical use.

7.3 Deformation-Aware Robotic US Imaging

In Chapter 4, we present a novel stiffness-based deformation correction method, incorporating image pixel displacements, contact forces, and nonlinear tissue stiffness, which is proposed to recover a zero-compression 3D tissue geometry from the deformed data recorded during robotic scans. Due to the inevitable pressure between probe and contact surface, the soft tissue's geometry is often distorted in B-mode images; thereby, such deformation hinders the consistent and accurate biometric measurement and diagnosis in clinical routines.

To this end, the proposed approach considered the nonlinear property of tissue stiffness as a key factor in correcting deformation. We obtained this patient-specified property by performing robotic palpation. Our approach enables the rapid adaptation of the optimized regression model to unseen positions by updating the local stiffness. Promising correction results have been achieved on both stiff and soft phantoms at arbitrary sampling positions. Additionally, the experimental results for 3D US acquired under different contact forces demonstrate that the proposed method is also able to recover zero-compression volumes from deformed images.

The pressure-induced deformation is a common issue for US imaging of soft tissues. With the proposed approach, accurate and reproducible 3D images, independent of the experience of sonographers, can be achieved by correcting the US deformation. To qualitatively and quantitatively validate the proposed approach, blood vessel is investigated as the target anatomy. But this study can benefit other applications requiring accurate geometrical measurements as well, such as examining and monitoring breast tumors [122] and soft tissue sarcomas [116]. Besides accurate 3D volumes, this study will also benefit the multimodal image fusion. The two typical clinical applications are image-guided intervention for soft tissues like breasts [123, 124], and imaging-guided orthopedic surgery [9]. For the former one, CT or MR is often used to provide high-resolution anatomies, while US images provide a live view during the intervention. Regarding the latter one, since patients may be moved after acquiring CT or MR images, it is necessary to do the registration between pre-operative images and patients by registering the live US images to the preoperative images. The corrected images could result in better results because it is easier to align the biological landmarks between the zero-compression B-mode images and preoperative images.

7.4 Autonomous Robotic US Screening of Tubular Structures

In Chapter 5, we present an end-to-end framework for automatic robotic US screening of tubular structures using only real-time US imaging feedback. The framework is inspired by the way sonographers conduct US scans. To this end, a neural network was trained to segment and track the tubular structures from cross-sectional US images during the screening process. To the best of our knowledge, this work is the first method able to automatically perform a US scan of a vascular structure with optimized probe position and orientation in real-time.

The proposed method can automatically adjust the probe position to horizontally center a tracked object and tune the probe orientation to the normal direction of the target vessel during US sweeps. Additionally, the radius of the target object is calculated in real-time, even when the probe is tilted from the normal direction. The results demonstrated that the proposed method is able to automatically perform US sweeps and accurately estimate the geometry of the target vessels both in simulation (e_{ra}: $1.16 \pm 0.1\ mm$, e_{or}: $2.7 \pm 3.3^{\circ}$) and for a real phantom (e_{ra}: $0.77 \pm 0.4\ mm$, e_{or}^{rea}: $3.7 \pm 1.6^{\circ}$, e_{ce}: $0.24 \pm 0.19\ mm$). The development of autonomous RUSS is a promising way to overcome inter-operator variability providing repeatable US images. In addition, with such a RUSS, sonographers can focus on diagnosis to fully utilize their experience and knowledge. This approach could be integrated with autonomous diagnosis techniques to further pave the way for a fully automatic US-guided intervention system [142].

7.5 Learning US Examination from Expert Demonstrations

In Chapter 6, we propose an advanced machine learning framework to explore standard scan planes defined by clinicians by demonstrations using robotic techniques. The ability to autonomously “explore" target anatomies and navigate a US probe to a relevant 2D plane by learning from expert demonstrations is important to further improve the level of intelligence of RUSS; thereby, further allowing to intuitively acquire high-level physiological knowledge from senior clinicians.

To address the challenges of inevitable sub-optimal US examination demonstrations, the self-supervised GPSR approach was used to infer the reward for individual US images by understanding the underlying clinicians’ intention. The ability to learn from demonstrations can contribute to the worldwide deployment of RUSS because it allows sonographers to intuitively transfer their high-level clinical knowledge to RUSS without any requirement for robotic programming skills. It is our hope that this would lead to more consistent and repeatable diagnosis being achievable among different sonographers, particularly for inexperienced operators.

The proposed approach has been validated both in a grid world and on physical gel phantoms, respectively. In the grid world, the experimental results clearly demonstrate that the proposed GPSR is superior to the state-of-the-art approaches (maximum entropy [155] and PTR [157]) to generate a closer reward map to the ground truth. In addition, the experimental results on two typical clinical tasks on a tabular phantom allowing slack along the centerline and challenging ex-vivo animal organ phantoms (chicken heart and lamb kidney) strictly requiring location accuracy are $6.2\pm1.0mm$, $4.3\pm1.8\ mm$ and $2.7\pm2.1\ mm$ in terms of e_d and $1.8\pm1.0^\circ$, $4.3 \pm 2.7^\circ$, and $5.2 \pm 2.9^\circ$ in terms of e_r, respectively. In our hope, the proposed intelligent robotic stenographer for discovering and locating standard scan planes can further enable the development of a fully automatic intervention system (i.e., vessels [138]) and extensive US examination programs for the early diagnosis and monitoring of internal lesions or tumor.

8

Open Challenges and Future Perspectives

8.1 Open Challenges

Medical robots have gained increased attention, in particular during the COVID-19 pandemic. The role of robotics in managing public health and infectious diseases has been widely discussed among the community [161, 162, 163, 164, 165]. In order to extensively apply RUSS, as well as other medical robots, in clinical routine, there are still some open challenges need to be further addressed. Here we highlighted 1) the acceptance of RUSS, and 2) ethical and legal issues.

8.1.1 Acceptance for Patients and Clinicians

The RUSS is designed to help both sonographers and patients in clinical routine. Besides demonstrating comparable or even better outcomes, the acceptance for RUSS is also important. Regarding the tele-sonography systems, Adams *et al.* indicated that all patients (18) were willing (89% were strongly willing and the remaining 11% were willing) to have another telerobotic examination [166]. A similar result was reported by [167], where 97% of 28 patients were willing to have another teleoperation scan. The numbers of involved patients in these two work are very limited. More comprehensive survey should be carried out in the future. In addition, to the best of the authors' knowledge, the attitudes for autonomous RUSS from the whole clinicians are stilling missing.

Since autonomous RUSS is partially or fully out of the control of sonographers, significant worries about safety arise, and further stress both patients and sonographers during the scanning. Thereby, the acceptance for autonomous RUSS may be less optimistic than it for teleoperation systems. To successfully deploy RUSS in clinical routines, acceptance from both patients and sonographers is important. To improve the acceptance of RUSS, robotic researchers should more closely work with clinicians.

8.1.2 Ethical and Legal Issues

The ethical and legal issues regarding medical robotics are still not clearly defined. The distribution of responsibility between sonographers and RUSS (or other surgical robotic systems) is still blurry. The missing of such regulations severely limit the clinical translation progress for such systems. Regarding the most representative surgical robotic systems, namely

Da Vinci, the responsibility is fully held by clinicians [27]. This is considered as the key characteristic leading to regulatory approval in the current era.

In order to properly address the ethical, regulatory, and legal issues for RUSS, Yang *et al.* divided surgical robotic into six subgroups in terms of autonomy levels: no autonomy, robot assistance, task autonomy, conditional autonomy, high autonomy, and full autonomy [168]. According to this classification method, commercial robots are still solidly resting at Level-0, while a much larger number of high-autonomy surgical robotic systems are waiting for clinical translation [27]. Since commercial surgical robotics is dominated by a few disproportionately large companies; thereby they have no rush in disrupting the status quo [27]. To establish such regulations for medical robots, O'Sullivan *et al.* defined three different responsibilities: (1) accountability: the capacity of a system to give an explanation for its actions; (2) liability: legal liability for potential damages caused by a robot; and (3) culpability: whom and how to implement punishment [169].

In addition, Vayena *et al.* discussed ethical and legal issues for digital health in terms of privacy and security, trust, and accountability [170]. As a large amount of data is often necessary for analysis, protecting privacy is undoubtedly important for avoiding misuse. Besides, public trust is of paramount importance. Vayena *et al.* considered that the creation of a culture of trust will enable all stakeholders to benefit from the development of digital health [170]. Similarly, Yang *et al.* summarised five increasingly pressing topics in terms of ethics for robotics and AI [171]. Besides aforementioned terms like responsibility, this works further emphasized some societal issues like potential influence on unemployment and human freedom. Due to the quick evolution of the area of medical robotics, a proper and comprehensive regulatory will boost a prosperous market and gradually benefit all stakeholders.

8.2 Future Perspectives

In the past two decades, the development of the medical robotic system was booming. Benefiting from the great contributions of the three different communities (the engineers, clinicians, and entrepreneurs), a few intelligent robotic systems (e.g., Da Vinci system) achieved great success both in economic and clinic fields. The existence of such successful intelligent systems really impresses both patients and clinicians in terms of the treatment performance; and thereby, the whole society gradually accepts the novel solutions. This positive movement will further motivate regulatory changes and societal perception of how healthcare should be offered [27]. Once a systematic regulation is established, engineers and entrepreneurs will polish their system to fulfill the needs of clinicians; and thus, bring a revolution of treatment approaches by providing additional cognitive and physical supports.

Since US examination is widely used for different organs in daily clinical routine and the imaging quality is suffered from the experience of sonographers, robot-assist US system is one of the most active and successful field of medical robotics. To overcome the limitations of conventional free-hand US examination (e.g., operator-dependent outcomes and well-trained operators), autonomous RUSS has gained increasing attention by researchers. The enabling techniques (force control, probe orientation optimization, and trajectory planning) have been well established; and some application-orientated advanced techniques (motion-aware and deformation-aware US imaging) gradually gain attention to further address a certain of practical challenges in real scenarios. However, the current developments are still not enough to guarantee the good performance of RUSS in diverse contexts. To further achieve this objective, three promising directions of ongoing and future researches are summarised as follows.

8.2.1 Specialized RUSS

Regrading RUSS, we consider the specialization can be represented by two different fields: 1) hardware systems including the robotic manipulator and US probe and 2) clinical applications. Most of the existing RUSS was developed based on commercial robotic arms and the two most common probes (linear or convex). These systems have already proved that the use of robots can bring a certain of merits, such as accurate control of probe position and pose and good reproducibility. As the enabling techniques have already been well developed, the development of the RUSS encounters bottlenecks in terms of safety and real clinical performance. To further address these issues, the development of specified RUSS to fulfill various clinical needs become necessary.

In contrast to commercial robotic manipulators, customized robotic systems for various clinical applications have the potential to improve safety, ease the complexity of the control task, improve examination efficiency, and achieve better results. In [172], Bao *et al.* designed a passive mechanism to hold a US probe using springs, which guarantees the patient's safety by limiting the maximum contact force. Such a specialized design of hardware effectively reduces the complexity of the controller because we don't need to fusion compliant force control in the centerline of the probe and the position control in other DOFs. Similarly, Lindenroth *et al.* presented a soft robotic end-effect for safe fetal imaging acquisition [173]. Besides, Housdes *et al.* proposed a 17-DOFs dual-probe system to efficiently and accurately for fetal imaging [174]. For autonomous image-guided vascular access, Chen *et al.* designed a compact robotic system to hold the US probe [138]. Besides these existing customized RUSS for different clinical applications, an even bold idea is to change the probe shape to better adapt the object surface, like a round probe for the scanning of limbs or breast.

8.2.2 Incorporating Additional Techniques to Realizing Advanced Functions

In real scenarios, only enabling techniques are often not enough to robustly and accurately accomplish specified US examinations. In real scenarios, only enabling techniques are usually not enough to robustly and accurately accomplish specified US examinations. For example, regarding the scanning of limbs arteries, the scan path is usually very long and the position of subjects is often changed by the sonographer to completely visualize the whole arteries [6]. Thus, this task is challenging because the robotic system only has limited working space and complex motion detection and compensation algorithm is necessary to be developed to continue the scan after the change of subject position [6]. By incorporating some computer vision techniques like marker identification and point cloud registration, a RUSS with an advanced function to monitor and compensate for the change of object pose was presented in [6]. Besides, Jiang *et al.* employed the optical flow technique to estimate the force-induced deformation field and further recovered the zero-pressure images from deformed images [5]. In addition, Patlan-Rosales *et al.* considered the RF signal acquired with different pressures to build a strain map; and then combined the technique of visual servoing to maintain the object in horizontally imaging center [175]. To provide more intuitive visualization, the augmented reality (AR) technique was employed by Shen *et al.* to visualize the identified tumors from intraoperative US imaging [176, 177].

8.2.3 Learning-Based RUSS

The development of AI techniques can benefit both imaging understanding and autonomous US probe manipulation. Regarding the former one, abundant studies have demonstrated that AI-based approaches achieved outperformance the traditional imaging process methods [178, 179, 180]. Benefiting from the accurate segmentation of target objects (e.g., blood vessels), an automatic US scan over a tubular structure and carotid based on real-time US images have been demonstrated in [7] and [181], respectively. The proper understanding of real-time US images enables proper navigation to automatically accomplish US examinations. To further allow RUSS to “explore" the standard US planes, RL has been employed [150, 182]. Besides, to allow senior sonographers to intuitively transfer their clinical knowledge and experience to RUSS, Burke *et al.* [157] and Mylonas *et al.* [156] proposed different frameworks to enable the ability of learning from demonstrations.

Although such systems have great potential to realize the high level of automation, the current systems still have great limitations. There are still some open challenges, such as how to adapt to the patient-specified parameter? how to reduce the training time? How to bridge the gap between the simulated environment and the real environment? In addition, the standards of the medical device prescribe a development process based on risk evaluation and minimization [27]. But the current AI-based approaches work like a black box, which prevents a detailed risk analysis. The development of the novel topic of explainable AI could be a solution for this issue in the future.

Part IV

Appendix

Abstracts of Publications not Discussed in this Dissertation

A

Towards Autonomous Atlas-Based Ultrasound Acquisitions in Presence of Articulated Motion [1]

Zhongliang Jiang*, Yuan Gao*, Le Xie, Nassir Navab

Robotic ultrasound (US) imaging aims at overcoming some of the limitations of free-hand US examinations, e.g. difficulty in guaranteeing intra- and inter-operator repeatability. However, due to anatomical and physiological variations between patients and relative movement of anatomical substructures, it is challenging to robustly generate optimal trajectories to examine the anatomies of interest, in particular, when they comprise articulated joints. To address this challenge, this paper proposes a vision-based approach allowing autonomous robotic US limb scanning. To this end, an atlas MRI template of a human arm with annotated vascular structures is used to generate trajectories and register and project them onto patients' skin surfaces for robotic US acquisition. To effectively segment and accurately reconstruct the targeted 3D vessel, we make use of spatial continuity in consecutive US frames by incorporating channel attention modules into a U-Net-type neural network. The automatic trajectory generation method is evaluated on six volunteers with various articulated joint angles. In all cases, the system can successfully acquire the planned vascular structure on volunteers' limbs. For one volunteer the MRI scan was also available, which allows the evaluation of the average radius of the scanned artery from US images, resulting in a radius estimation (1.2±0.05 mm) comparable to the MRI ground truth (1.2±0.04 mm).

IEEE Robotics and Automation Letters(2022)

VesNet-RL: Simulation-based Reinforcement Learning for Real-World US Probe Navigation [2]

Yuan Bi*, **Zhongliang Jiang***, Yuan Gao, Thomas Wendler, Angelos Karlas, Nassir Navab

Ultrasound (US) is one of the most common medical imaging modalities since it is radiation-free, low-cost, and real-time. In freehand US examinations, sonographers often navigate a US probe to visualize standard examination planes with rich diagnostic information. However, reproducibility and stability of the resulting images often suffer from intra- and inter-operator variation. Reinforcement learning (RL), as an interaction-based learning method, has demonstrated its effectiveness in visual navigating tasks; however, RL is limited in terms of generalization. To address this challenge, we propose a simulation-based RL framework for real-world navigation of US probes towards the standard longitudinal views of vessels. A UNet is used to provide binary masks from US images; thereby, the RL agent trained on simulated binary vessel images can be applied in real scenarios without further training. To accurately characterize actual states, a multi-modality state representation structure is introduced to facilitate the understanding of environments. Moreover, considering the characteristics of vessels, a novel standard view recognition approach based on the minimum bounding rectangle is proposed to terminate the searching process. To evaluate the effectiveness of the proposed method, the trained policy is validated virtually on 3D volumes of a volunteer's in-vivo carotid artery, and physically on custom-designed gel phantoms using robotic US. The results demonstrate that proposed approach can effectively and accurately navigate the probe towards the longitudinal view of vessels.

IEEE Robotics and Automation Letters(2022)

Deep Reinforcement Learning Based Trajectory Planning Under Uncertain Constraints [4]

Lienhung Chen, **Zhongliang Jiang**, Long Cheng, Alois C Knoll, Mingchuan Zhou

With the advance in algorithms, deep reinforcement learning (DRL) offers solutions to trajectory planning under uncertain environments. Different from traditional trajectory planning which requires lots of effort to tackle complicated high-dimensional problems, the recently proposed DRL enables the robot manipulator to autonomously learn and discover optimal trajectory planning by interacting with the environment. In this article, we present state-of-the-art DRL-based collision-avoidance trajectory planning for uncertain environments such as a safe human coexistent environment. Since the robot manipulator operates in high dimensional continuous state-action spaces, model-free, policy gradient-based soft actor-critic (SAC), and deep deterministic policy gradient (DDPG) framework are adapted to our scenario for comparison. In order to assess our proposal, we simulate a 7-DOF Panda (Franka Emika) robot manipulator in the PyBullet physics engine and then evaluate its trajectory planning with reward, loss, safe rate, and accuracy. Finally, our final report shows the effectiveness of state-of-the-art DRL algorithms for trajectory planning under uncertain environments with zero collision after 5,000 episodes of training.

Frontiers in Neurorobotics (2022)

Diffeomorphic Respiratory Motion Estimation of Thoracoabdominal Organs for Image-Guided Interventions [8]

Long Lei, Li Huang, Baoliang Zhao, Ying Hu, **Zhongliang Jiang**, Jianwei Zhang, Bing Li

Purpose: *Percutaneous image-guided interventions are commonly used for the diagnosis and treatment of cancer. In practice, physiological breathing-induced motion increases the difficulty of accurately inserting needles into tumors without impairing the surrounding vital structures. In this work, we propose a data-driven patient-specific hierarchical respiratory motion estimation framework to accurately estimate the position of a tumor and surrounding vital tissues in real time.* ***Methods:*** *The motion of optical markers attached to the chest or abdomen skin is used as a surrogate signal to estimate tumor motion based on* ϵ*-support vector regression (*ϵ*-SVR). With the estimated tumor motion as the input, a novel respiratory motion model is developed to estimate the diffeomorphic deformation field of the whole organ (liver or lung) without intraoperative, iterative optimization computations. The respiratory motion model of the whole organ is established in Lie algebra space based on the kriging algorithm to ensure that the estimated deformation field is diffeomorphic, optimal, and unbiased. Preoperative prior knowledge for modeling the motion of whole organs is obtained by deformation registration between four-dimensional computed tomography (4D CT) images using a hybrid diffeomorphic registration method.* ***Results and Conclusions:*** *Experimental results on an in vivo beagle dog show that the minimum value of the determinant of the Jacobian of the estimated deformation field is greater than zero, so the estimated deformation field of the whole liver with our method is diffeomorphic. The mean position error of the tumor is 1.2 mm corresponding to a mean accuracy improvement of 76.5%, and the mean position error of the whole liver is 2.1 mm, corresponding to a mean accuracy improvement of 37.9%. The experimental results based on public human subject data show that the mean position error of the tumor is 1.1 mm, corresponding to a mean accuracy improvement of 83.1%, and the mean position error of the whole lung is 2.1 mm, corresponding to a mean accuracy improvement of 41.4%. The positioning errors for the tumor and whole organ are hierarchical and consistent with clinical demand.*

Medical Physics (2021)

Model-Based Compensation of Moving Tissue for State Recognition in Robotic-Assisted Pedicle Drilling [10]

Zhongliang Jiang, Long Lei, Yu Sun, Xiaozhi Qi, Ying Hu, Bing Li, Nassir Navab, Jianwei Zhang

Drilling is one of the hardest parts of pedicle screw fixation, and it is one of the most dangerous operations because inaccurate screw placement would injury vital tissues, particularly when the vertebra is not stationary. Here we demonstrate the drilling state recognition method for moving tissue by compensating the displacement based on a simplified motion predication model of a vertebra with respect to the tidal volume. To adapt it to different patients, the prediction model was built based on the physiological data recorded from subjects themselves. In addition, the spindle speed of the drilling tool was investigated to find a suitable speed for the robotic-assisted system. To ensure patient safety, a monitoring system was built based on the thrusting force and tracked position information. Finally, experiments were carried out on a fresh porcine lamellar bone fixed on a 3-PRS parallel robot used to simulate the vertebra displacement. The success rate of the robotic-assisted drilling procedure reached 95% when the moving bone was compensated.

IEEE Transactions on Medical Robotics and Bionics (2020)

State Recognition of Decompressive Laminectomy with Multiple Information in Robot-Assisted Surgery [12]

Yu Sun, Li Wang, **Zhongliang Jiang**, Bing Li, Ying Hu, Wei Tian

The decompressive laminectomy is a common operation for treatment of lumbar spinal stenosis. The tools for grinding and drilling are used for fenestration and internal fixation, respectively. The state recognition is one of the main technologies in robot-assisted surgery, especially in tele-surgery, because surgeons have limited perception during remote-controlled robot-assisted surgery. The novelty of this paper is that a state recognition system is proposed for the robot-assisted tele-surgery. By combining the learning methods and traditional methods, the robot from the slave-end can think about the current operation state like a surgeon, and provide more information and decision suggestions to the master-end surgeon, which aids surgeons work safer in tele-surgery. For the fenestration, we propose an image-based state recognition method that consists a U-Net derived network, grayscale redistribution and dynamic receptive field assisting in controlling the grinding process to prevent the grinding-bit from crossing the inner edge of the lamina to damage the spinal nerves. For the internal fixation, we propose an audio and force-based state recognition method that consists signal features extraction methods, LSTM-based prediction and information fusion assisting in monitoring the drilling process to prevent the drilling-bit from crossing the outer edge of the vertebral pedicle to damage the spinal nerves. Several experiments are conducted to show the reliability of the proposed system in robot-assisted surgery.

Artificial Intelligence in Medicine (2020)

Safety Control Method of Robot-Assisted Cataract Surgery with Virtual Fixture and Virtual Force Feedback [13]

Yongfei Yang, **Zhongliang Jiang**, Yuanyuan Yang, Xiaozhi Qi, Ying Hu, Jianjun Du, Bing Han, Guiqin Liu

Surgery is an effective means of treating cataracts and restoring vision. However, cataract surgery rate (CSR) in developing countries and regions is relatively low due to the lack of experienced high-level surgeons. In this paper, to reduce the reliance of surgery on physician experience and thereby increase CSR, a master-slave robotic system and safety control strategies with a virtual fixture and virtual force feedback are proposed to assist cataract surgery. First, the surgery is divided into four different stages with different robot control modes. Secondly, the virtual constraint area with virtual spring model in the operating stage is established, so that the doctor can distinguish the operation area where the end of the surgical instrument is located by feedback force. Thirdly, safety control algorithm guarantees that the surgical instrument strictly moves around the surgical incision point, which is regarded as a remote centre of motion, so that the cornea outside the incision point is not injured. Finally, the experimental results show that the proposed safety control strategy allows the robotic system to perform the procedure safely.

Journal of Intelligent & Robotic Systems (2020)

Cutting Depth Monitoring based on Milling Force for Robot-Assisted Laminectomy [14]

Zhongliang Jiang, Xiaozhi Qi, Yu Sun, Ying Hu, Guillaume Zahnd, Jianwei Zhang

Goal: *In the context of robot-assisted laminectomy surgery, an analytical force model is introduced to guarantee procedural safety. The aim of the method is to intraoperatively monitor the cutting depth via modeling the milling status.* ***Methods:*** *The theoretical dynamic model for the surgical milling process is based on the flute geometry of the ball-end milling tool. A particle swarm optimization algorithm is exploited to calibrate the model using the local average force, and to validate it using the denoised dynamic force. A wear detection method based on the fast Fourier transform is proposed to determine the quality of the tool geometry and to avoid using worn tools, which may lead to imprecise and unsafe operations.* ***Results:*** *Milling experiments were performed on machined fresh bovine femur bones. The experimental results thus obtained from the mechanical model are in good accordance with the numerical model. The proposed method can monitor the current cutting depth with an accuracy of ±0.1 mm in regions located within the depth [0.8-1.2 mm], and ±0.2 mm within [1.2-1.6 mm].* ***Conclusion:*** *The proposed model can successfully estimate the milling force and the cutting depth intraoperatively in experimental conditions. Significance: This approach has the potential to improve the safety of laminectomy operations in humans, and make it more accessible to younger surgeons by lowering the required manual skills threshold.*

IEEE Transactions on Automation Science and Engineering (2019)

Model-based Spinal Deformation Compensation in Robot-Assisted Decompressive Laminectomy [15]

Yu Sun, **Zhongliang Jiang**, Xiaozhi Qi, Ying Hu, Bing Li, Jianwei Zhang

Decompressive laminectomy is a common operation for the treatment of lumbar spinal stenosis. A high-speed burr and piezosurgery are often used to remove the lamina to relieve compressed nerves. However, surgeons need to control the cutting margin of the lamina based on experience, and the remaining laminar thickness is difficult to quantify before image measurements. In other surgeries, robot-assisted grinding guided by navigation can make operations accurate, but the deformation of bone tissue caused by instrument interaction cannot be ignored, especially for spinal surgery. Because of the elasticity of the intervertebral disk, the spine deforms under external force, which leads to deviation from the preoperative trajectory for the robot. This paper proposes a model-based compensation method for spinal deformation during decompression operation. A torsion model and bending model are built based on traditional elastic mechanics, and then a coupling model is simplified and modified for improved robot control. The source and influence of residual error caused by the noise is analysed, and data-based estimation of the model parameters is conducted to accommodate different patients. Robot deformation is also considered in the control system, and the model inputs—the feedback force data—are processed using the adaptive Kalman filter in real time. The lumbar data are segmented for finite element analysis, which provides the input and output data with which the parameters of the model are estimated, and the simulation shows that the signal-to-noise ratio of the force sensor has a great effect on system performance.

Mechatronics (2019)

Robot-Assisted Decompressive Laminectomy Planning based on 3D Medical Image [16]

Yu Sun, **Zhongliang Jiang**, Xiaozhi Qi, Ying Hu, Bing Li, Jianwei Zhang

The decompressive laminectomy is a common treatment for lumbar spinal stenosis. Generally, surgeons use grinding tools to remove laminae under the guidance of intraoperative medical images. To improve accuracy and reduce surgeons' burdens, robot-assisted surgery is gaining acceptance. This paper proposes a method to plan grinding paths and velocities based on 3-D medical images in the context of robot-assisted decompressive laminectomies. As the lesion areas to be grinded are irregular, an interactive method with 3-D reconstruction is designed for surgeons to transfer discrete information about grinding paths and velocities to the robot system. The path generation strategy is based on a ray casting algorithm after space registrations, while the velocity generation strategy is based on the virtual force and mechanical analysis is used to optimize temporal efficiency and stability. A complete system is developed to test and explore the feasibility of this method. Results suggest that robot-assisted decompressive laminectomies can be performed well.

IEEE Access (2018)

A Model of Vertebral Motion and Key Point Recognition of Drilling with Force in Robot-Assisted Spinal Surgery [17]

Zhongliang Jiang, Yu Sun, Shijia Zhao, Ying Hu, Jianwei Zhang

Pedicle drilling is a crucial and high-risk process in spinal surgery. Due to the respiration and cardiac cycle, the position of spine would fluctuate during operations, which result in an increase of the difficulty in state recognition of pedicle drilling. To guarantee the safety and validity, a model-based compensation method is proposed in this paper. To build the empirical model of vertebral motion, vertebral displacement and tidal volume (Tv) signals are collected from volunteers. To rule out disturbances in original signal, FFT and wavelet transform (DWT) are used to process the experimental signals. In order to select the apt basis for different signals, the root mean square error decision-making method (RMSE-DMM) is introduced. When the filtered vertebral displacement signal is obtained, the particle swarm optimization (PSO) algorithm is used to figure out the empirical model. The robot assisted systems (RAS) can easily compensate vertebral fluctuation based on the empirical model. Due to the goal of pedicle drilling is to drill a hole from surface of first cortical layer to the interior of second cortical layer, a new key point recognition algorithm, based on force, proposed in this paper. To verify the effectiveness of compensation and the recognition algorithm, 3 sets of comparison experiments are carried out. And the results of experiment show the compensation method and new key point recognition algorithm perform effectively.

2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2017)

Bibliography

[1] Z. Jiang, Y. Gao, L. Xie, and N. Navab. "Towards Autonomous Atlas-Based Ultrasound Acquisitions in Presence of Articulated Motion". In: *IEEE Robotics and Automation Letters* (2022) (cit. on pp. xvii, 153).

[2] 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* (2022) (cit. on pp. xvii, 154).

[3] Z. Jiang, N. Danis, Y. Bi, et al. "Precise Repositioning of Robotic Ultrasound: Improving Registration-Based Motion Compensation Using Ultrasound Confidence Optimization". In: *IEEE Transactions on Instrumentation and Measurement* 71 (2022), pp. 1–11 (cit. on pp. xvii, 52).

[4] L. Chen, Z. Jiang, L. Cheng, A. C. Knoll, and M. Zhou. "Deep Reinforcement Learning Based Trajectory Planning Under Uncertain Constraints". In: *Frontiers in Neurorobotics* 16 (2022) (cit. on pp. xvii, 155).

[5] 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 pp. xvii, 12, 54, 78, 82, 85, 87, 89, 92, 94, 96, 150).

[6] 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 pp. xvii, 12, 52, 57, 58, 61, 63–66, 69, 73, 79, 83, 91, 117, 149, 150).

[7] Z. Jiang, Z. Li, M. Grimm, et al. "Autonomous robotic screening of tubular structures based only on real-time ultrasound imaging feedback". In: *IEEE Transactions on Industrial Electronics* (2021) (cit. on pp. xvii, 12, 65, 67, 73, 83, 100, 101, 103, 106, 108, 110–113, 116, 137, 150).

[8] L. Lei, L. Huang, B. Zhao, et al. "Diffeomorphic respiratory motion estimation of thoracoabdominal organs for image-guided interventions". In: *Medical Physics* 48.8 (2021), pp. 4160–4176 (cit. on pp. xviii, 156).

[9] Z. Jiang, M. Grimm, M. Zhou, Y. Hu, J. Esteban, and N. Navab. "Automatic force-based probe positioning for precise robotic ultrasound acquisition". In: *IEEE Transactions on Industrial Electronics* (2020) (cit. on pp. xviii, 12, 18, 19, 21–23, 37, 41–43, 45, 47, 49, 57, 63, 83, 96, 101, 109, 113, 120, 145).

[10] Z. Jiang, L. Lei, Y. Sun, et al. "Model-based compensation of moving tissue for state recognition in robotic-assisted pedicle drilling". In: *IEEE Transactions on Medical Robotics and Bionics* 2.3 (2020), pp. 463–473 (cit. on pp. xviii, 65, 157).

[11] Z. Jiang, M. Grimm, M. Zhou, et al. "Automatic normal positioning of robotic ultrasound probe based only on confidence map optimization and force measurement". In: *IEEE Robotics and Automation Letters* 5.2 (2020), pp. 1342–1349 (cit. on pp. xviii, 18, 21, 24, 25, 28–30, 32–36, 43, 45, 46, 48, 49, 54, 59, 70, 83, 96, 101, 109, 113, 120).

[12] Y. Sun, L. Wang, Z. Jiang, B. Li, Y. Hu, and W. Tian. “State recognition of decompressive laminectomy with multiple information in robot-assisted surgery”. In: *Artificial Intelligence in Medicine* 102 (2020), p. 101763 (cit. on pp. xviii, 158).

[13] Y. Yang, Z. Jiang, Y. Yang, et al. “Safety control method of robot-assisted cataract surgery with virtual fixture and virtual force feedback”. In: *Journal of Intelligent & Robotic Systems* 97.1 (2020), pp. 17–32 (cit. on pp. xviii, 159).

[14] Z. Jiang, X. Qi, Y. Sun, Y. Hu, G. Zahnd, and J. Zhang. “Cutting depth monitoring based on milling force for robot-assisted laminectomy”. In: *IEEE Transactions on Automation Science and Engineering* 17.1 (2019), pp. 2–14 (cit. on pp. xviii, 20, 42, 160).

[15] Y. Sun, Z. Jiang, X. Qi, Y. Hu, B. Li, and J. Zhang. “Model-based spinal deformation compensation in robot-assisted decompressive laminectomy”. In: *Mechatronics* 59 (2019), pp. 115–126 (cit. on pp. xviii, 161).

[16] Y. Sun, Z. Jiang, X. Qi, Y. Hu, B. Li, and J. Zhang. “Robot-assisted decompressive laminectomy planning based on 3D medical image”. In: *IEEE Access* 6 (2018), pp. 22557–22569 (cit. on pp. xviii, 57, 162).

[17] Z. Jiang, Y. Sun, S. Zhao, Y. Hu, and J. Zhang. “A model of vertebral motion and key point recognition of drilling with force in robot-assisted spinal surgery”. In: *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2017, pp. 6455–6462 (cit. on pp. xix, 65, 163).

[18] T. L. Szabo. *Diagnostic ultrasound imaging: inside out*. Academic press, 2004 (cit. on p. 3).

[19] P. R. Hoskins, K. Martin, and A. Thrush. *Diagnostic ultrasound: physics and equipment*. CRC Press, 2019 (cit. on pp. 3, 6, 7, 9, 10, 19, 119).

[20] 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).

[21] G. Brown. “Work related musculoskeletal disorders in sonographers”. In: *BMUS Bulletin* 11.3 (2003), pp. 6–13 (cit. on p. 3).

[22] L. Chan, T. Fung, T. Leung, D. Sahota, and T. Lau. “Volumetric (3D) imaging reduces inter- and intraobserver variation of fetal biometry measurements”. In: *Ultrasound in Obstetrics and Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology* 33.4 (2009), pp. 447–452 (cit. on pp. 3, 121).

[23] J. P. Kösters and P. C. Gøtzsche. “Regular self-examination or clinical examination for early detection of breast cancer”. In: *Cochrane Database of Systematic Reviews* 2 (2003) (cit. on p. 4).

[24] P. P. Sengupta, N. Narula, K. Modesto, et al. “Feasibility of intercity and trans-Atlantic telerobotic remote ultrasound: assessment facilitated by a nondedicated bandwidth connection”. In: *JACC: Cardiovascular Imaging* 7.8 (2014), pp. 804–809 (cit. on p. 4).

[25] M. Georgescu, A. Sacccomandi, B. Baudron, and P. L. Arbeille. “Remote sonography in routine clinical practice between two isolated medical centers and the university hospital using a robotic arm: a 1-year study”. In: *Telemedicine and e-Health* 22.4 (2016), pp. 276–281 (cit. on p. 4).

[26] *Intuitive Surgical I Da Vinci systems*. `https://www.intuitive.com/en-us/healthcare-professionals/surgeons`. Accessed: 2022-01-28 (cit. on p. 4).

[27] A. Attanasio, B. Scaglioni, E. De Momi, P. Fiorini, and P. Valdastri. “Autonomy in surgical robotics”. In: *Annual Review of Control, Robotics, and Autonomous Systems* 4 (2021), pp. 651–679 (cit. on pp. 4, 148, 150).

[28] P. Arbeille, A. Capri, J. Ayoub, V. Kieffer, M. Georgescu, and G. Poisson. “Use of a robotic arm to perform remote abdominal telesonography”. In: *American journal of Roentgenology* 188.4 (2007), W317–W322 (cit. on p. 4).

[29] P. Sekar and V. Vilvanathan. "Telecardiology: effective means of delivering cardiac care to rural children". In: *Asian Cardiovascular and Thoracic Annals* 15.4 (2007), pp. 320–323 (cit. on p. 4).

[30] H. Q. Lean and Y. Zhou. "Acoustic field of phased-array ultrasound transducer with the focus/foci shifting". In: *Journal of Medical and Biological Engineering* 39.6 (2019), pp. 919–931 (cit. on p. 7).

[31] Q. Zhou, K. H. Lam, H. Zheng, W. Qiu, and K. K. Shung. "Piezoelectric single crystal ultrasonic transducers for biomedical applications". In: *Progress in materials science* 66 (2014), pp. 87–111 (cit. on p. 8).

[32] R. Tsumura, J. W. Hardin, K. Bimbraw, et al. "Tele-Operative Low-Cost Robotic Lung Ultrasound Scanning Platform for Triage of COVID-19 Patients". In: *IEEE Robotics and Automation Letters* 6.3 (2021), pp. 4664–4671 (cit. on pp. 11, 12).

[33] R. Ye, X. Zhou, F. Shao, et al. "Feasibility of a 5G-based robot-assisted remote ultrasound system for cardiopulmonary assessment of patients with Coronavirus Disease 2019". In: *Chest* 159.1 (2021), pp. 270–281 (cit. on p. 11).

[34] K. Masuda, E. Kimura, N. Tateishi, and K. Ishihara. "Three dimensional motion mechanism of ultrasound probe and its application for tele-echography system". In: *Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No. 01CH37180)*. Vol. 2. IEEE. 2001, pp. 1112–1116 (cit. on p. 11).

[35] M. Mitsuishi, S. Warisawa, T. Tsuda, et al. "Remote ultrasound diagnostic system". In: *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No. 01CH37164)*. Vol. 2. IEEE. 2001, pp. 1567–1574 (cit. on p. 11).

[36] P. Abolmaesumi, S. E. Salcudean, W.-H. Zhu, S. P. DiMaio, and M. R. Sirouspour. "A user interface for robot-assisted diagnostic ultrasound". In: *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No. 01CH37164)*. Vol. 2. IEEE. 2001, pp. 1549–1554 (cit. on p. 11).

[37] 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. 11).

[38] R. Tsumura and H. Iwata. "Robotic fetal ultrasonography platform with a passive scan mechanism". In: *International journal of computer assisted radiology and surgery* 15.8 (2020), pp. 1323–1333 (cit. on pp. 11, 12).

[39] N. Koizumi, S. Warisawa, M. Nagoshi, H. Hashizume, and M. Mitsuishi. "Construction methodology for a remote ultrasound diagnostic system". In: *IEEE Transactions on Robotics* 25.3 (2009), pp. 522–538 (cit. on p. 11).

[40] K. Ito, S. Sugano, and H. Iwata. "Portable and attachable tele-echography robot system: FASTele". In: *2010 annual international conference of the ieee engineering in medicine and biology*. IEEE. 2010, pp. 487–490 (cit. on p. 11).

[41] A. Filippeschi, F. Brizzi, E. Ruffaldi, J. M. Jacinto Villegas, L. Landolfi, and C. A. Avizzano. "Evaluation of diagnostician user interface aspects in a virtual reality-based tele-ultrasonography simulation". In: *Advanced Robotics* 33.15-16 (2019), pp. 840–852 (cit. on p. 11).

[42] M. W. Gilbertson and B. W. Anthony. "Force and position control system for freehand ultrasound". In: *IEEE Trans. Robot.* 31.4 (2015), pp. 835–849 (cit. on pp. 12, 18, 54, 83, 101).

[43] W.-H. Zhu, S. E. Salcudean, S. Bachmann, and P. Abolmaesumi. "Motion/force/image control of a diagnostic ultrasound robot". In: *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065)*. Vol. 2. IEEE. 2000, pp. 1580–1585 (cit. on p. 12).

[44] 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 pp. 12, 121).

[45] F. Pierrot, E. Dombre, E. Dégoulange, et al. “Hippocrate: A safe robot arm for medical applications with force feedback”. In: *Med. Image Anal.* 3.3 (1999), pp. 285–300 (cit. on pp. 12, 101, 120).

[46] G. Ma, S. R. Oca, Y. Zhu, P. J. Codd, and D. M. Buckland. “A Novel Robotic System for Ultrasound-guided Peripheral Vascular Localization”. In: *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2021, pp. 12321–12327 (cit. on p. 12).

[47] F. Suligoj, C. M. Heunis, J. Sikorski, and S. Misra. “RobUSt–An Autonomous Robotic Ultrasound System for Medical Imaging”. In: *IEEE Access* 9 (2021), pp. 67456–67465 (cit. on p. 12).

[48] 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 (2016), pp. 538–548 (cit. on pp. 12, 13, 56, 66, 67, 84, 86, 106, 110, 132).

[49] M. Welleweerd, A. de Groot, V Groenhuis, F. Siepel, and S Stramigioli. “Out-of-Plane Corrections for Autonomous Robotic Breast Ultrasound Acquisitions”. In: *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2021, pp. 12515–12521 (cit. on p. 12).

[50] 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* 12.6 (2017), pp. 983–991 (cit. on p. 12).

[51] G. Ning, X. Zhang, and H. Liao. “Autonomic Robotic Ultrasound Imaging System Based on Reinforcement Learning”. In: *IEEE Transactions on Biomedical Engineering* (2021) (cit. on p. 12).

[52] G. Ning, J. Chen, X. Zhang, and H. Liao. “Force-guided autonomous robotic ultrasound scanning control method for soft uncertain environment”. In: *International Journal of Computer Assisted Radiology and Surgery* (2021), pp. 1–11 (cit. on p. 12).

[53] 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 pp. 12, 54, 70, 101, 121).

[54] S. Wang, R. J. Housden, Y. Noh, et al. “Analysis of a customized clutch joint designed for the safety management of an ultrasound robot”. In: *Applied Sciences* 9.9 (2019), p. 1900 (cit. on p. 12).

[55] B. Ihnatsenka and A. P. Boezaart. “Ultrasound: Basic understanding and learning the language”. In: *Int. J. Shoulder Surg.* 4.3 (2010), p. 55 (cit. on pp. 12, 19, 59, 70, 109).

[56] Q. Huang, J. Lan, and X. Li. “Robotic arm based automatic ultrasound scanning for three-dimensional imaging”. In: *IEEE Trans. Ind. Inform.* 15.2 (2018), pp. 1173–1182 (cit. on pp. 12, 13, 20, 50, 54, 83, 101, 109, 121, 143).

[57] 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. 12, 48, 54, 59, 128).

[58] X. Ma, Z. Zhang, and H. K. Zhang. “Autonomous Scanning Target Localization for Robotic Lung Ultrasound Imaging”. In: *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2021, pp. 9467–9474 (cit. on p. 12).

[59] P. Chatelain, A. Krupa, and N. Navab. “Optimization of ultrasound image quality via visual servoing”. In: *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*. IEEE. 2015, pp. 5997–6002 (cit. on pp. 12, 20).

[60] P. Chatelain, A. Krupa, and N. Navab. “Confidence-driven control of an ultrasound probe: Target-specific acoustic window optimization”. In: *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*. IEEE. 2016, pp. 3441–3446 (cit. on pp. 12, 101).

[61] A. Karamalis, W. Wein, T. Klein, and N. Navab. “Ultrasound confidence maps using random walks”. In: *Med. Image Anal.* 16.6 (2012), pp. 1101–1112 (cit. on pp. 12, 17, 23, 26, 70).

[62] P. Chatelain, A. Krupa, and N. Navab. “Confidence-driven control of an ultrasound probe”. In: *IEEE Trans. Robot.* 33.6 (2017), pp. 1410–1424 (cit. on pp. 12, 19, 27, 70).

[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 pp. 13, 53).

[64] M. Akbari, J. Carriere, R. Sloboda, et al. “Robot-assisted Breast Ultrasound Scanning Using Geometrical Analysis of the Seroma and Image Segmentation”. In: *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 3784–3791 (cit. on p. 13).

[65] 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 pp. 13, 101).

[66] 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. 13).

[67] A. A. Moshaii and F. Najafi. “A review of robotic mechanisms for ultrasound examinations”. In: *Industrial Robot: An International Journal* (2014) (cit. on p. 13).

[68] R. Elek, T. D. Nagy, D. A. Nagy, et al. “Robotic platforms for ultrasound diagnostics and treatment”. In: *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE. 2017, pp. 1752–1757 (cit. on p. 13).

[69] 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. 13).

[70] L. Nouaille, M. A. Laribi, C. A. Nelson, S Zeghloul, and G. Poisson. “Review of kinematics for minimally invasive surgery and tele-echography robots”. In: *Journal of Medical Devices* 11.4 (2017) (cit. on p. 13).

[71] 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* (2021), pp. 1–17 (cit. on p. 13).

[72] 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* (2021) (cit. on p. 13).

[73] M. R. Bonyadi and Z. Michalewicz. “Particle swarm optimization for single objective continuous space problems: a review”. In: *Evolutionary computation* 25.1 (2017), pp. 1–54 (cit. on p. 17).

[74] C. M. Court-Brown and B. Caesar. “Epidemiology of adult fractures: a review”. In: *Injury* 37.8 (2006), pp. 691–697 (cit. on p. 18).

[75] M. Ciganovic, F. Ozdemir, F. Pean, P. Fuernstahl, C. Tanner, and O. Goksel. “Registration of 3D freehand ultrasound to a bone model for orthopedic procedures of the forearm”. In: *International journal of computer assisted radiology and surgery* 13.6 (2018), pp. 827–836 (cit. on p. 18).

[76] W. Wein, A. Karamalis, A. Baumgartner, and N. Navab. “Automatic bone detection and soft tissue aware ultrasound–CT registration for computer-aided orthopedic surgery”. In: *International Journal of Computer Assisted Radiology and Surgery* 10.6 (2015), pp. 971–979 (cit. on p. 18).

[77] M. Salehi, R. Prevost, J.-L. Moctezuma, N. Navab, and W. Wein. “Precise ultrasound bone registration with learning-based segmentation and speed of sound calibration”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2017, pp. 682–690 (cit. on pp. 18, 21).

[78] Y. Han, L. Zhang, M. Guo, C. Fan, and F. Liang. “Tool paths generation strategy for polishing of freeform surface with physically uniform coverage”. In: *The International Journal of Advanced Manufacturing Technology* 95.5 (2018), pp. 2125–2144 (cit. on p. 19).

[79] L. Zhang, Y. Han, C. Fan, Y. Tang, and X. Song. “Polishing path planning for physically uniform overlap of polishing ribbons on freeform surface”. In: *The International Journal of Advanced Manufacturing Technology* 92.9 (2017), pp. 4525–4541 (cit. on p. 19).

[80] F. Tian, Z. Li, C. Lv, and G. Liu. “Polishing pressure investigations of robot automatic polishing on curved surfaces”. In: *The International Journal of Advanced Manufacturing Technology* 87.1 (2016), pp. 639–646 (cit. on p. 19).

[81] J.-P. Merlet. “C-surface applied to the design of an hybrid force-position robot controller”. In: *Proceedings. 1987 IEEE International Conference on Robotics and Automation*. Vol. 4. IEEE. 1987, pp. 1055–1059 (cit. on pp. 19, 48).

[82] P. Kazanzides, N. S. Bradley, and W. A. Wolovich. “Dual-drive force/velocity control: implementation and experimental results”. In: *1989 IEEE International Conference on Robotics and Automation*. IEEE Computer Society. 1989, pp. 92–93 (cit. on pp. 19, 48).

[83] T. Yoshikawa and A. Sudou. “Dynamic hybrid position/force control of robot manipulators-on-line estimation of unknown constraint”. In: *IEEE Transactions on Robotics and Automation* 9.2 (1993), pp. 220–226 (cit. on p. 19).

[84] J. J. Kaufman and T. A. Einhorn. “Ultrasound assessment of bone”. In: *Journal of Bone and Mineral Research* 8.5 (1993) (cit. on p. 21).

[85] O. Zettinig, B. Frisch, S. Virga, et al. “3D ultrasound registration-based visual servoing for neurosurgical navigation”. In: *International journal of computer assisted radiology and surgery* 12.9 (2017), pp. 1607–1619 (cit. on pp. 22, 41).

[86] G. Welch, G. Bishop, et al. “An introduction to the Kalman filter”. In: (1995) (cit. on p. 28).

[87] K Choi, J. L. Kuhn, M. J. Ciarelli, and S. A. Goldstein. “The elastic moduli of human subchondral, trabecular, and cortical bone tissue and the size-dependency of cortical bone modulus”. In: *Journal of biomechanics* 23.11 (1990), pp. 1103–1113 (cit. on p. 36).

[88] J. Kennedy and R. Eberhart. “Particle swarm optimization”. In: *Proceedings of ICNN’95-international conference on neural networks*. Vol. 4. IEEE. 1995, pp. 1942–1948 (cit. on p. 42).

[89] R. G. Lyons. “Understanding Digital Signal Processing, (2004)”. In: *Motoyama T.(2005), Study on Variable Resolution Imaging in a Microscope OCT System, null* 336 (2004) (cit. on p. 43).

[90] D. L. Longo, A. S. Fauci, D. L. Kasper, S. L. Hauser, J Jamenson, et al. *Harrinson´ s manual of medicine*. 2013 (cit. on p. 52).

[91] G. J. Hankey, P. E. Norman, and J. W. Eikelboom. “Medical treatment of peripheral arterial disease”. In: *Jama* 295.5 (2006), pp. 547–553 (cit. on pp. 52, 100).

[92] F. G. R. Fowkes, D. Rudan, I. Rudan, et al. “Comparison of global estimates of prevalence and risk factors for peripheral artery disease in 2000 and 2010: a systematic review and analysis”. In: *The lancet* 382.9901 (2013), pp. 1329–1340 (cit. on p. 53).

[93] R. Stein, I. Hriljac, J. L. Halperin, S. M. Gustavson, V. Teodorescu, and J. W. Olin. “Limitation of the resting ankle-brachial index in symptomatic patients with peripheral arterial disease”. In: *Vascular medicine* 11.1 (2006), pp. 29–33 (cit. on pp. 53, 100).

[94] V. Aboyans, J.-B. Ricco, M.-L. E. Bartelink, M. Björck, M. Brodmann, T. Cohnert, et al. "2017 ESC Guidelines on the Diagnosis and Treatment of Peripheral Arterial Diseases, in collaboration with the European Society for Vascular Surgery (ESVS)". In: *European heart journal* 39.9 (2018), pp. 763–816 (cit. on p. 53).

[95] R Collins, G Cranny, J Burch, et al. "A systematic review of duplex ultrasound, magnetic resonance angiography and computed tomography angiography for the diagnosis and assessment of symptomatic, lower limb peripheral arterial disease." In: *Health Technology Assessment (Winchester, England)* 11.20 (2007), pp. iii–184 (cit. on p. 53).

[96] A Fenster, C Blake, I. Gyacskov, A Landry, and J. Spence. "3D ultrasound analysis of carotid plaque volume and surface morphology". In: *Ultrasonics* 44 (2006), e153–e157 (cit. on p. 53).

[97] J. A. Hossack, T. S. Sumanaweera, S. Napel, and J. S. Ha. "Quantitative 3-D diagnostic ultrasound imaging using a modified transducer array and an automated image tracking technique". In: *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 49.8 (2002), pp. 1029–1038 (cit. on p. 53).

[98] A. Gee, R. Prager, G. Treece, and L. Berman. "Engineering a freehand 3D ultrasound system". In: *Pattern Recognition Letters* 24.4-5 (2003), pp. 757–777 (cit. on p. 54).

[99] 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 pp. 54, 104, 117).

[100] S.-Y. Sun, B. W. Anthony, and M. W. Gilbertson. "Trajectory-based deformation correction in ultrasound images". In: *Medical Imaging 2010: Ultrasonic Imaging, Tomography, and Therapy*. Vol. 7629. International Society for Optics and Photonics. 2010, 76290A (cit. on pp. 54, 83, 87, 88, 92, 93).

[101] S. Garrido-Jurado, R. Munoz-Salinas, F. J. Madrid-Cuevas, and R. Medina-Carnicer. "Generation of fiducial marker dictionaries using mixed integer linear programming". In: *Pattern Recognition* 51 (2016), pp. 481–491 (cit. on p. 57).

[102] R. F. Rahmat, T. Chairunnisa, D. Gunawan, and O. S. Sitompul. "Skin color segmentation using multi-color space threshold". In: *2016 3rd International Conference on Computer and Information Sciences (ICCOINS)*. IEEE. 2016, pp. 391–396 (cit. on p. 58).

[103] 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: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2016, pp. 388–396 (cit. on pp. 65, 78, 101).

[104] C. Balakrishna, S. Dadashzadeh, and S. Soltaninejad. "Automatic detection of lumen and media in the IVUS images using U-Net with VGG16 Encoder". In: *arXiv preprint arXiv:1806.07554* (2018) (cit. on pp. 67, 72).

[105] D. Dunbar and G. Humphreys. "A spatial data structure for fast Poisson-disk sample generation". In: *ACM Transactions on Graphics (TOG)* 25.3 (2006), pp. 503–508 (cit. on p. 69).

[106] D. Holz, A. E. Ichim, F. Tombari, R. B. Rusu, and S. Behnke. "Registration with the point cloud library: A modular framework for aligning in 3-D". In: *IEEE Robotics & Automation Magazine* 22.4 (2015), pp. 110–124 (cit. on pp. 69, 74).

[107] A. A. Pravitasari, N. Iriawan, M. Almuhayar, et al. "UNet-VGG16 with transfer learning for MRI-based brain tumor segmentation". In: *Telkomnika* 18.3 (2020), pp. 1310–1318 (cit. on p. 72).

[108] O. Ronneberger, P. Fischer, and T. Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241 (cit. on pp. 72, 102).

[109] K. Simonyan and A. Zisserman. "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556* (2014) (cit. on p. 72).

[110] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. “Imagenet: A large-scale hierarchical image database”. In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee. 2009, pp. 248–255 (cit. on p. 72).

[111] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman. “LabelMe: a database and web-based tool for image annotation”. In: *International journal of computer vision* 77.1-3 (2008), pp. 157–173 (cit. on p. 72).

[112] D. P. Kingma and J. Ba. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014) (cit. on pp. 74, 90, 110, 135).

[113] Z. Min, D. Zhu, H. Ren, and M. Q.-H. Meng. “Feature-guided nonrigid 3-D point set registration framework for image-guided liver surgery: From isotropic positional noise to anisotropic positional noise”. In: *IEEE Transactions on Automation Science and Engineering* 18.2 (2020), pp. 471–483 (cit. on p. 78).

[114] Z. Min, J. Wang, J. Pan, and M. Q.-H. Meng. “Generalized 3-D point set registration with hybrid mixture models for computer-assisted orthopedic surgery: From isotropic to anisotropic positional error”. In: *IEEE Transactions on Automation Science and Engineering* (2020) (cit. on p. 78).

[115] K. Pfister, W. Schierling, E. M. Jung, H. Apfelbeck, C. Hennersperger, and P. M. Kasprzak. “Standardized 2D ultrasound versus 3D/4D ultrasound and image fusion for measurement of aortic aneurysm diameter in follow-up after EVAR”. In: *Clinical hemorheology and microcirculation* 62.3 (2016), pp. 249–260 (cit. on p. 83).

[116] S. Virga, R. Göbl, M. Baust, N. Navab, and C. Hennersperger. “Use the force: Deformation correction in robotic 3D ultrasound”. In: *International journal of computer assisted radiology and surgery* 13.5 (2018), pp. 619–627 (cit. on pp. 83, 86, 87, 95, 97, 145).

[117] G. M. Treece, R. W. Prager, A. H. Gee, and L Berman. “Correction of probe pressure artifacts in freehand 3D ultrasound”. In: *Med. Image Anal.* 6.3 (2002), pp. 199–214 (cit. on pp. 83, 87).

[118] M. R. Burcher, L. Han, and J. A. Noble. “Deformation correction in ultrasound images using contact force measurements”. In: *Proceedings IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA 2001)*. IEEE. 2001, pp. 63–70 (cit. on p. 83).

[119] B. Flack, M. Makhinya, and O. Goksel. “Model-based compensation of tissue deformation during data acquisition for interpolative ultrasound simulation”. In: *2016 IEEE International Symposium on Biomedical Imaging (ISBI)*. IEEE. 2016, pp. 502–505 (cit. on p. 83).

[120] J. Dahmani, Y. Petit, and C. Laporte. “Model-based correction of ultrasound image deformations due to probe pressure”. In: *Medical Imaging 2017: Image Processing*. Vol. 10133. International Society for Optics and Photonics. 2017, p. 101331D (cit. on p. 83).

[121] B. D. Lucas, T. Kanade, et al. “An iterative image registration technique with an application to stereo vision”. In: *Proc. Int. Joint Conf. Artif. Intell*. 1981, pp. 674–679 (cit. on p. 88).

[122] T. S. Pheiffer, R. C. Thompson, D. C. Rucker, A. L. Simpson, and M. I. Miga. “Model-based correction of tissue compression for tracked ultrasound in soft tissue image-guided surgery”. In: *Ultrasound in medicine & biology* 40.4 (2014), pp. 788–803 (cit. on pp. 95, 145).

[123] T. S. Pheiffer and M. I. Miga. “Toward a generic real-time compression correction framework for tracked ultrasound”. In: *Int. J. Comput. Assist. Radiol. Surg.* 10.11 (2015), pp. 1777–1792 (cit. on pp. 96, 145).

[124] E. Tagliabue, D. Dall’Alba, E. Magnabosco, C. Tenga, I. Peterlík, and P. Fiorini. “Position-based modeling of lesion displacement in ultrasound-guided breast biopsy”. In: *International journal of computer assisted radiology and surgery* 14.8 (2019), pp. 1329–1339 (cit. on pp. 96, 145).

[125] R. Kojcev, B. Fuerst, O. Zettinig, et al. “On the reproducibility of expert-operated and robotic ultrasound acquisitions”. In: *Int. J. Comput. Assist. Radiol. Surg.* 12.6 (2017), pp. 1003–1011 (cit. on p. 97).

[126] T. H. Wong, K. H. Tay, M. G. Sebastian, and S. G. Tan. “Duplex ultrasonography arteriography as first-line investigation for peripheral vascular disease”. In: *Singap. Med. J.* 54.5 (2013), pp. 271–274 (cit. on p. 100).

[127] M. Davis, M. Harris, and J. J. Earnshaw. “Implementation of the national health service abdominal aortic aneurysm screening program in England”. In: *J. Vasc. Surg.* 57.5 (2013), pp. 1440–1445 (cit. on p. 100).

[128] J. H. Yoon, M. H. Kim, E.-K. Kim, H. J. Moon, J. Y. Kwak, and M. J. Kim. “Interobserver variability of ultrasound elastography: how it affects the diagnosis of breast lesions”. In: *American Journal of Roentgenology* 196.3 (2011), pp. 730–736 (cit. on p. 100).

[129] F. Conti, J. Park, and O. Khatib. “Interface design and control strategies for a robot assisted ultrasonic examination system”. In: *Exp. Robot.* Springer. 2014, pp. 97–113 (cit. on p. 101).

[130] M. L. Balter, A. I. Chen, T. J. Maguire, and M. L. Yarmush. “Adaptive kinematic control of a robotic venipuncture device based on stereo vision, ultrasound, and force guidance”. In: *IEEE Trans. Ind. Electron.* 64.2 (2016), pp. 1626–1635 (cit. on p. 101).

[131] D. Mishra, S. Chaudhury, M. Sarkar, and A. S. Soin. “Ultrasound image segmentation: a deeply supervised network with attention to boundaries”. In: *IEEE Trans. Biomed. Eng.* 66.6 (2018), pp. 1637–1648 (cit. on pp. 102, 122).

[132] J. Guerrero, S. E. Salcudean, J. A. McEwen, and et. al. “Real-time vessel segmentation and tracking for ultrasound imaging applications”. In: *IEEE Trans. Med. Imag.* 26.8 (2007), pp. 1079–1090 (cit. on pp. 102, 107).

[133] E. Smistad and F. Lindseth. “Real-Time automatic artery segmentation, reconstruction and registration for ultrasound-guided regional anaesthesia of the femoral nerve”. In: *IEEE Trans. Med. Imag.* 35.3 (2015), pp. 752–761 (cit. on pp. 102, 107).

[134] A. Crimi, M. Makhinya, U. Baumann, and et. al. “Automatic measurement of venous pressure using B-mode ultrasound”. In: *IEEE Trans. Biomed. Eng.* 63.2 (2015), pp. 288–299 (cit. on pp. 102, 107).

[135] E. Karami, M. S. Shehata, and A. Smith. “Adaptive polar active contour for segmentation and tracking in ultrasound videos”. In: *IEEE Trans. Circuits Syst. Video Technol.* 29.4 (2018), pp. 1209–1222 (cit. on p. 102).

[136] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever. “Multiscale vessel enhancement filtering”. In: *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv. (MICCAI)*. Springer. 1998, pp. 130–7 (cit. on p. 102).

[137] G. Litjens, T. Kooi, B. E. Bejnordi, et al. “A survey on deep learning in medical image analysis”. In: *Med. Image Anal.* 42 (2017), pp. 60–88 (cit. on p. 102).

[138] A. I. Chen, M. L. Balter, T. J. Maguire, and M. L. Yarmush. “Deep learning robotic guidance for autonomous vascular access”. In: *Nat. Mach. Intell.* 2.2 (2020), pp. 104–115 (cit. on pp. 102, 120, 121, 140, 146, 149).

[139] Y. Wu and K. He. “Group normalization”. In: *Proc. Eur. Conf. comput. vis. (ECCV)*. 2018, pp. 3–19 (cit. on p. 104).

[140] S. S. Mao, N. Ahmadi, B. Shah, et al. “Normal thoracic aorta diameter on cardiac computed tomography in healthy asymptomatic adults: impact of age and gender”. In: *Academic radiology* 15.7 (2008), pp. 827–834 (cit. on p. 112).

[141] H. Patel and I. Hacihaliloglu. “Improved Automatic Bone Segmentation Using Large-Scale Simulated Ultrasound Data to Segment Real Ultrasound Bone Surface Data”. In: *Proc. IEEE Int. Conf. Bioinform. Bioeng. (BIBE)*. IEEE. 2020, pp. 288–294 (cit. on p. 117).

[142] J. Guo, Y. Liu, and et al. “A novel robotic guidance system with eye gaze tracking control for needle based interventions”. In: *IEEE Trans. Cogn. Dev. Syst.* 13.1 (2019), pp. 179–188 (cit. on pp. 117, 145).

[143] F. Y. Rizi, J. Au, H. Yli-Ollila, et al. “Carotid wall longitudinal motion in ultrasound imaging: An expert consensus review”. In: *Ultrasound in Medicine & Biology* (2020) (cit. on pp. 120, 121).

[144] C. F. Baumgartner, K. Kamnitsas, J. Matthew, et al. “SonoNet: real-time detection and localisation of fetal standard scan planes in freehand ultrasound”. In: *IEEE transactions on medical imaging* 36.11 (2017), pp. 2204–2215 (cit. on pp. 120, 121).

[145] M. B. Stone, C. Moon, D. Sutijono, and M. Blaivas. “Needle tip visualization during ultrasound-guided vascular access: short-axis vs long-axis approach”. In: *The American journal of emergency medicine* 28.3 (2010), pp. 343–347 (cit. on p. 120).

[146] M. A. Maraci, R. Napolitano, A. Papageorghiou, and J. A. Noble. “Searching for structures of interest in an ultrasound video sequence”. In: *International Workshop on Machine Learning in Medical Imaging*. Springer. 2014, pp. 133–140 (cit. on p. 121).

[147] H. Chen, D. Ni, J. Qin, et al. “Standard plane localization in fetal ultrasound via domain transferred deep neural networks”. In: *IEEE journal of biomedical and health informatics* 19.5 (2015), pp. 1627–1636 (cit. on p. 121).

[148] R. Droste, L. Drukker, A. T. Papageorghiou, and J. A. Noble. “Automatic Probe Movement Guidance for Freehand Obstetric Ultrasound”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2020, pp. 583–592 (cit. on p. 121).

[149] H. Hase, M. F. Azampour, M. Tirindelli, et al. “Ultrasound-guided robotic navigation with deep reinforcement learning”. In: *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS)*. 2020, pp. 5534–5541 (cit. on pp. 121, 123).

[150] K. Li, J. Wang, Y. Xu, et al. “Autonomous Navigation of an Ultrasound Probe Towards Standard Scan Planes with Deep Reinforcement Learning”. In: *arXiv preprint arXiv:2103.00718* (2021) (cit. on pp. 122, 123, 150).

[151] A. Kurenkov, A. Mandlekar, R. Martin-Martin, S. Savarese, and A. Garg. “AC-Teach: A Bayesian Actor-Critic Method for Policy Learning with an Ensemble of Suboptimal Teachers”. In: *Conference on Robot Learning*. PMLR. 2020, pp. 717–734 (cit. on p. 122).

[152] S. Ross, G. Gordon, and D. Bagnell. “A reduction of imitation learning and structured prediction to no-regret online learning”. In: *Proceedings of the fourteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings. 2011, pp. 627–635 (cit. on p. 122).

[153] P. Abbeel and A. Y. Ng. “Apprenticeship learning via inverse reinforcement learning”. In: *Proceedings of the twenty-first international conference on Machine learning*. 2004, p. 1 (cit. on p. 122).

[154] B. D. Ziebart, A. L. Maas, J. A. Bagnell, and A. K. Dey. “Maximum entropy inverse reinforcement learning.” In: *Aaai*. Vol. 8. Chicago, IL, USA. 2008, pp. 1433–1438 (cit. on p. 122).

[155] N. Aghasadeghi and T. Bretl. “Maximum entropy inverse reinforcement learning in continuous state spaces with path integrals”. In: *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2011, pp. 1561–1566 (cit. on pp. 122, 132, 133, 139, 146).

[156] 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 pp. 122, 150).

[157] M. Burke, K. Lu, D. Angelov, et al. “Learning robotic ultrasound scanning using probabilistic temporal ranking”. In: *arXiv preprint arXiv:2002.01240* (2020) (cit. on pp. 122–124, 128, 132, 133, 135, 138, 139, 146, 150).

[158] C. Eteke, D. Kebüde, and B. Akgün. “Reward Learning From Very Few Demonstrations”. In: *IEEE Transactions on Robotics* (2020) (cit. on p. 123).

[159] D. P. Kingma and M. Welling. “Auto-encoding variational bayes”. In: *arXiv preprint arXiv:1312.6114* (2013) (cit. on pp. 123, 125).

[160] A. Lasso, T. Heffter, A. Rankin, C. Pinter, T. Ungi, and G. Fichtinger. “PLUS: open-source toolkit for ultrasound-guided intervention systems”. In: *IEEE transactions on biomedical engineering* 61.10 (2014), pp. 2527–2537 (cit. on p. 129).

[161] G. Hager, V. Kumar, R. Murphy, D. Rus, and R. Taylor. “The role of robotics in infectious disease crises”. In: *arXiv preprint arXiv:2010.09909* (2020) (cit. on p. 147).

[162] A. Zemmar, A. M. Lozano, and B. J. Nelson. “The rise of robots in surgical environments during COVID-19”. In: *Nature Machine Intelligence* 2.10 (2020), pp. 566–572 (cit. on p. 147).

[163] A. Di Lallo, R. Murphy, A. Krieger, J. Zhu, R. H. Taylor, and H. Su. “Medical Robots for Infectious Diseases: Lessons and Challenges from the COVID-19 Pandemic”. In: *IEEE Robotics & Automation Magazine* 28.1 (2021), pp. 18–27 (cit. on p. 147).

[164] G.-Z. Yang, B. J. Nelson, R. R. Murphy, et al. *Combating COVID-19—The role of robotics in managing public health and infectious diseases*. 2020 (cit. on p. 147).

[165] A. Gao, R. R. Murphy, W. Chen, et al. “Progress in robotics for combating infectious diseases”. In: *Science Robotics* 6.52 (2021) (cit. on p. 147).

[166] S. J. Adams, B. E. Burbridge, A. Badea, et al. “Initial experience using a telerobotic ultrasound system for adult abdominal sonography”. In: *Canadian Association of Radiologists’ Journal* 68.3 (2017), pp. 308–314 (cit. on p. 147).

[167] S. J. Adams, B. E. Burbridge, A. Badea, et al. “A crossover comparison of standard and telerobotic approaches to prenatal sonography”. In: *Journal of Ultrasound in Medicine* 37.11 (2018), pp. 2603–2612 (cit. on p. 147).

[168] G.-Z. Yang, J. Cambias, K. Cleary, et al. “Medical robotics—Regulatory, ethical, and legal considerations for increasing levels of autonomy”. In: *Science Robotics* 2.4 (2017), p. 8638 (cit. on p. 148).

[169] S. O’Sullivan, N. Nevejans, C. Allen, et al. “Legal, regulatory, and ethical frameworks for development of standards in artificial intelligence (AI) and autonomous robotic surgery”. In: *The International Journal of Medical Robotics and Computer Assisted Surgery* 15.1 (2019), e1968 (cit. on p. 148).

[170] E. Vayena, T. Haeusermann, A. Adjekum, and A. Blasimme. “Digital health: meeting the ethical and policy challenges”. In: *Swiss medical weekly* 148 (2018), w14571 (cit. on p. 148).

[171] G.-Z. Yang, J. Bellingham, P. E. Dupont, et al. “The grand challenges of Science Robotics”. In: *Science robotics* 3.14 (2018), eaar7650 (cit. on p. 148).

[172] X. Bao, S. Wang, R Housden, J. Hajnal, and K. Rhode. “A Constant-Force End-Effector With Online Force Adjustment for Robotic Ultrasonography”. In: *IEEE robotics and automation letters* 6.2 (2021), pp. 2547–2554 (cit. on p. 149).

[173] L. Lindenroth, R. J. Housden, S. Wang, J. Back, K. Rhode, and H. Liu. “Design and integration of a parallel, soft robotic end-effector for extracorporeal ultrasound”. In: *IEEE Transactions on Biomedical Engineering* (2019) (cit. on p. 149).

[174] J. Housden, S. Wang, X. Bao, et al. “Towards Standardized Acquisition with a Dual-probe Ultrasound Robot for Fetal Imaging”. In: *IEEE robotics and automation letters* 6.2 (2021), pp. 1059–1065 (cit. on p. 149).

[175] P. A. Patlan-Rosales and A. Krupa. “Strain estimation of moving tissue based on automatic motion compensation by ultrasound visual servoing”. In: *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2017, pp. 2941–2946 (cit. on p. 150).

[176] J. Shen, N. Zemiti, C. Taoum, et al. “Transrectal ultrasound image-based real-time augmented reality guidance in robot-assisted laparoscopic rectal surgery: a proof-of-concept study”. In: *International journal of computer assisted radiology and surgery* 15.3 (2020), pp. 531–543 (cit. on p. 150).

[177] G. Samei, K. Tsang, C. Kesch, et al. “A partial augmented reality system with live ultrasound and registered preoperative MRI for guiding robot-assisted radical prostatectomy”. In: *Medical image analysis* 60 (2020), p. 101588 (cit. on p. 150).

[178] V. Bass, J. Mateos, I. M. Rosado-Mendez, and J. Márquez. “Ultrasound image segmentation methods: A review”. In: *AIP Conference Proceedings*. Vol. 2348. 1. AIP Publishing LLC. 2021, p. 050018 (cit. on p. 150).

[179] M. H. Hesamian, W. Jia, X. He, and P. Kennedy. “Deep learning techniques for medical image segmentation: achievements and challenges”. In: *Journal of digital imaging* 32.4 (2019), pp. 582–596 (cit. on p. 150).

[180] C. Shan, T. Tan, J. Han, and D. Huang. “Ultrasound tissue classification: a review”. In: *Artificial Intelligence Review* 54.4 (2021), pp. 3055–3088 (cit. on p. 150).

[181] Y. Huang, W. Xiao, C. Wang, H. Liu, R. Huang, and Z. Sun. “Towards Fully Autonomous Ultrasound Scanning Robot With Imitation Learning Based on Clinical Protocols”. In: *IEEE Robotics and Automation Letters* 6.2 (2021), pp. 3671–3678 (cit. on p. 150).

[182] Y. Chen, C. Yang, M. Li, and S.-L. Dai. “Learning to Predict Action Based on B-ultrasound Image Information”. In: *2021 6th IEEE International Conference on Advanced Robotics and Mechatronics (ICARM)*. IEEE. 2021, pp. 492–497 (cit. on p. 150).

List of Figures

List of Tables