Clermont Auvergne University
Engineering Sciences Doctoral School
Pascal Institute / CHU Clermont-Ferrand

Thesis

for the degree of

Doctor of Science

Presented and defended by

Mohammad Mahdi KALANTARI

Multimodal Registration for Augmented Reality Guidance in Laparoscopic Liver Resection: the Contribution of Laparoscopic Ultrasound

Ph.D. Directors:	Adrien BARTOLI	-	Université Clermont Auvergne and CHU Clermont-Ferrand
	Youcef MEZOUAR	-	Clermont Auvergne INP
Co-supervision:	Erol OZGUR	-	Clermont Auvergne INP
	Mohammad ALKHATIB	-	Clermont Auvergne INP

Defended on July 9th, 2026

Jury Members:

Reviewers:	François CHADEBECQ	-	Middlesex University London
	Omar TAHRI	-	Université Bourgogne Europe
Examiners:	Caroline PETITJEAN	-	Université de Rouen

Abstract:

Laparoscopic liver resection (LLR) offers substantial clinical benefits over open surgery, including reduced morbidity, shorter hospitalization, and faster recovery. However, its broader adoption remains constrained by a fundamental limitation of minimally invasive surgery (MIS): the loss of direct palpation and restricted visualization of subsurface anatomy. In LLR, tumors and critical vascular structures are often hidden beneath the liver surface, forcing surgeons to rely on mental registration between preoperative imaging and sparse intraoperative cues. This cognitive burden is further exacerbated by liver deformation, limited depth perception, and the fragmented use of multimodal imaging. This thesis addresses these challenges by introducing a multimodal registration and augmented reality (AR) framework for intraoperative guidance in LLR, centered on the integration of laparoscopic ultrasound (LUS) as an intraoperative source of subsurface anatomical information.

The proposed framework is built upon three technical contributions. First, a novel marker-free and sensor-free method is introduced for real-time estimation of the 5-degree-of-freedom (DoF) pose of a LUS probe from monocular laparoscopic images, with 5-DoF being the maximum recoverable pose owing to the probe's geometric symmetry. By modeling the probe head as a spherocylinder and exploiting its geometric constraints, the method enables single-shot pose recovery without external tracking, markers, or manual initialization. Second, the ambiguity inherent to monocular spherocylindrical pose estimation is resolved through complementary strategies for recovering the sixth DoF, combining shaft-based constraints, motion priors, and temporal filtering. This yields robust and temporally stable 6-DoF probe tracking and enables real-time augmentation of the ultrasound imaging plane directly within the laparoscopic view. Third, the estimated probe pose and LUS-derived tumor observations are integrated into a multimodal registration framework that aligns preoperative CT/MRI models with intraoperative laparoscopic and ultrasound data, enabling coherent full-tumor augmentation in situ.

The proposed methods are validated through a comprehensive experimental protocol spanning clinical, ex vivo, and phantom datasets, specifically curated to capture the geometric and procedural variability of laparoscopic liver surgery. This includes more than 3,000 annotated clinical images, controlled ex vivo acquisitions with 6-DoF ground truth, and deformable phantom studies with known geometric correspondences. Across these settings, the framework demonstrates robust real-time performance for probe pose estimation and AR, in addition to clinically relevant multimodal registration accuracy. Collectively, this work establishes a complete and practical pipeline for sensorless surgical modalities guidance in laparoscopic liver resection and provides a foundation for future research in image-guided MIS.

Keywords: laparoscopic liver resection, augmented reality, laparoscopic ultrasound, multimodal registration, surgical navigation, image-guided surgery.

Résumé :

La résection hépatique laparoscopique (RHL) offre des avantages cliniques substantiels par rapport à la chirurgie ouverte, notamment une réduction de la morbidité, une hospitalisation plus courte et une récupération plus rapide. Cependant, son adoption à plus grande échelle reste entravée par une limitation fondamentale de la chirurgie mini-invasive : la perte de la palpation directe et une visualisation restreinte de l'anatomie sous-parenchymateuse. En RHL, les tumeurs et les structures vasculaires critiques sont souvent dissimulées sous la surface du foie, contraignant les chirurgiens à s'appuyer sur un recalage mental entre l'imagerie préopératoire et des indices peropératoires parcellaires. Cette charge cognitive est accentuée par la déformation du foie, une perception limitée de la profondeur et l'utilisation fragmentée de l'imagerie multimodale. Cette thèse répond à ces défis en introduisant un cadre de recalage multimodal et de réalité augmentée (RA) pour le guidage peropératoire en RHL, centré sur l'intégration de l'échographie laparoscopique (EUL) comme source d'information anatomique sous-corticale.

Le cadre proposé repose sur trois contributions techniques. Premièrement, une méthode inédite, sans marqueur ni capteur, est introduite pour l'estimation en temps réel de la pose à 5 degrés de liberté (ddl) d'une sonde d'échographie laparoscopique à partir d'images laparoscopiques monoculaires, les 5 ddl constituant la pose maximale pouvant être estimée en raison de la symétrie géométrique de la sonde. En modélisant la tête de la sonde comme un sphérocylindre et en exploitant ses contraintes géométriques, cette méthode permet de déterminer la pose à partir d'une seule image (single-shot) sans suivi externe, sans marqueur, ni initialisation manuelle. Deuxièmement, l'ambiguïté inhérente à l'estimation monoculaire de la pose d'un sphérocylindre est résolue par des stratégies complémentaires pour récupérer le sixième degré de liberté, combinant des contraintes basées sur la tige de la sonde, des modèles de mouvement (priors) et un filtrage temporel. Cela permet d'obtenir un suivi de la sonde à 6-ddl robuste et temporellement stable, autorisant l'augmentation en temps réel du plan d'imagerie échographique directement dans la vue laparoscopique. Troisièmement, la pose estimée de la sonde et les observations tumorales issues de l'EUL sont intégrées dans un cadre de recalage multimodal qui aligne les modèles CT/IRM préopératoires avec les données laparoscopiques et échographiques peropératoires, permettant ainsi une augmentation cohérente de la tumeur complète in situ.

Les méthodes proposées sont validées par un protocole expérimental complet comprenant des données cliniques, ex vivo, et des fantômes, spécifiquement sélectionnés pour capturer la variabilité géométrique et procédurale de la chirurgie hépatique laparoscopique. Cela inclut plus de 3 000 images cliniques annotées, des acquisitions ex vivo contrôlées avec vérité terrain à 6-ddl, et des études sur fantômes déformables avec correspondances géométriques connues. Dans ces différents contextes, le protocole démontre des performances robustes en temps réel pour l'estimation de la position de la sonde et RA et une précision de recalage multimodal cliniquement pertinente. Collectivement, ces travaux établissent un pipeline complet et pratique pour le guidage chirurgical multimodal sans capteur en résection hépatique laparoscopique et fournissent une base pour les recherches futures en chirurgie mini-invasive guidée par l'image.

Mots-clés : résection hépatique laparoscopique, réalité augmentée, échographie laparoscopique, recalage multimodal, navigation chirurgicale, chirurgie guidée par l'image.

Valorization:

- **International Conferences**
  - **Markerless Ultrasound Probe Pose Estimation in Mini-Invasive Surgery**
    M. M. Kalantari, E. Ozgur, M. Alkhatib, E. Buc, B. Le Roy, R. Modrzejewski, Y. Mezouar, A. Bartoli
    2024 IEEE International Conference on Robotics and Automation (ICRA)
  - **Laparoscopic Augmented Reality from Laparoscopic Ultrasound**
    M. M. Kalantari, E. Ozgur, M. Alkhatib, E. Buc, B. Le Roy, R. Modrzejewski, Y. Mezouar, A. Bartoli
    2024 International Conference on Information Processing in Computer-Assisted Interventions (IPCAI)
  - **Stronger Together: Registering Preoperative Imagery, LUS, and MIS Liver Images**
    M. M. Kalantari, E. Ozgur, M. Alkhatib, N. Rabbani, Y. Espinel, R. Modrzejewski, B. Le Roy, E. Buc, Y. Mezouar, A. Bartoli
    2025 International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)
    [Selected for Spotlight Presentation]

- **Journals**
  - **Laparoscopic Augmented Reality from Laparoscopic Ultrasound**
    M. M. Kalantari, E. Ozgur, M. Alkhatib, E. Buc, B. Le Roy, R. Modrzejewski, Y. Mezouar, A. Bartoli
    International Journal of Computer Assisted Radiology and Surgery (IJCARS), special issue: selected papers from IPCAI 2024, July 2024

- **Data Challenges**
  - **Ovarian Cancer with Peritoneal Carcinomatosis**
    Participated as part of the *AURAI* team
    PINKCC Data Challenge — Finalist (June 2025)
  - **Detection of a Pancreatic Mass**
    Participated as part of the laboratory's team
    JFR Data Challenge — 4th Place (October 2023)

- **Supervision**
  - **Master's Intern:** Karima Cherfioui
    Topic: Development of A Vision-based Ultrasound Probe Pose Estimation Pipeline
    Duration: Apr 2025 – Sept 2025
- **Reviewer for International Conferences**
  - 2025 International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)

Acknowledgments

Many individuals have supported, guided, and inspired me throughout the different stages of my life, which have ultimately led me here. To briefly acknowledge those who were directly involved in this journey, I would first like to express my sincere gratitude to Adrien Bartoli for his insightful guidance, thoughtful coordination, and unwavering support at every stage of this work. I am also thankful to Youcef Mezouar, Erol Ozgur, and Mohammad Alkhatib for enriching the supervisory team and for sharing their valuable expertise and perspectives. Emmanuel Buc and Bertrand Le Roy for sharing their valuable clinical feedback. My appreciation also extends to SURGAR for providing the intellectual resources and technical support that greatly facilitated this research. I would additionally like to thank Daniel Pizarro for his valuable remarks and constructive feedback.

I am equally grateful to my previous supervisors at various stages of my academic and professional development, Hamid Taghirad, Ahmad Khalilpour, Philippe Cardou, and Mohammad Manthouri, whose mentorship laid much of the foundation upon which this work was built. I also wish to express my deep appreciation to Babak Seyfe, Sajjad Mahdizadeh, and Yaser Emamy, whose inspiration and perspectives have profoundly shaped my vision and approach.

Special acknowledgment goes to Navid Rabbani, whose role in developing methodologies and designing experiments was pivotal to this work; to Mathieu Labrunie, for generously sharing his knowledge and expertise; and to Karima Cherfioui, for her direct collaboration in the project. I am also thankful to colleagues who became much more than collaborators, Rasoul Sharifian, Yamid Espinel, Mina Zamani, Ivan Mikhailov, Benoit Magnin, Mickael Mathieu, Emilien Gadoux, Maria Bamaarouf, Anestis Christidis, and Devesh Adlakha, as well as Daniel González, Hubert Villeneuve, and Mélodie Daniel, all of whom contributed to making this journey both professionally rewarding and personally meaningful.

A heartfelt thank you goes to Navid, Neda, Rasoul, and Mina for being wonderful friends and making life here significantly easier and more joyful. I also warmly acknowledge Françoise and Ludivine, along with my artist friends Sarah, Fatemeh, and Delphine, whose presence and friendship brought beauty and balance to my life beyond academia.

My deepest gratitude goes to my longtime friends Yasin, Mahdi, Ali, Mohammad Mahdi, Sajjad, and Roya, for their constant emotional support and companionship, and to Sina, for always being there for me.

Finally, I would like to express my sincere gratitude to my beloved sister, Zeinab, and to my family for their support and encouragement. Their love and trust have been a great source of strength throughout this journey.

Contents

List of Acronyms

Chapter 1

Introduction

Contents

1.1 Context and Motivation

In this section, we first discuss the broader context of Minimally Invasive Surgery (MIS). We then focus specifically on Laparoscopic Liver Resection (LLR), highlighting the existing clinical and technological limitations in current practice. Building upon this analysis, we demonstrate how the *IMMORTALLS* project is positioned to address these gaps by introducing advanced intraoperative guidance and multimodal integration strategies.

1.1.1 Minimally Invasive Surgery (MIS)

In a discipline defined by continuous evolution in methodology and technology, the transition from open procedures to MIS represents the most profound paradigm shift in the surgical landscape over the last four decades [Fox 2025]. Across diverse surgical subspecialties, including general surgery, neurosurgery, and urology, the adoption of MIS for procedures such as cholecystectomies, craniotomies, and prostatectomies has consistently demonstrated clinical efficacy. MIS frequently yields patient outcomes comparable to or superior to those of traditional open approaches, while reducing postoperative complications, accelerating recovery, and shortening hospital stay [Hale *et al.* 2025, Xiao *et al.* 2024, Jaffray 2005]. For example, among 19,769 patients in colorectal surgery, hospital stay days are reduced to 5.6 for robotic MIS in comparison to 11.2 for open surgery [Vazquez *et al.* 2025] and the blood transfusion is reduced by 75% in a study over more than 1,000,000 patients across multiple oncologic specialties [Ricciardi *et al.* 2024].

Despite these clinical advantages, the implementation of MIS entails significant trade-offs. Primary among these are the fiscal burdens associated with specialized infrastructure [Fox 2026] and the high cost of recurring consumables [Al Farai *et al.* 2024]. Furthermore, MIS is frequently characterized by prolonged operative and setup durations [Lai *et al.* 2024]. Beyond logistical concerns, the inherent lack of haptic feedback presents a formidable clinical hurdle, particularly in intricate oncological resections where tactile perception is vital for tissue differentiation [Friebe 2026].

To contextualize the distinct operational demands of each approach, Fig. 1.1 provides a visual comparison between conventional Open Liver Resection (OLR) and a MIS LLR. This comparison highlights not only the stark difference in incision size and surgical access but also underscores the disparate visual fields available to the surgical team.

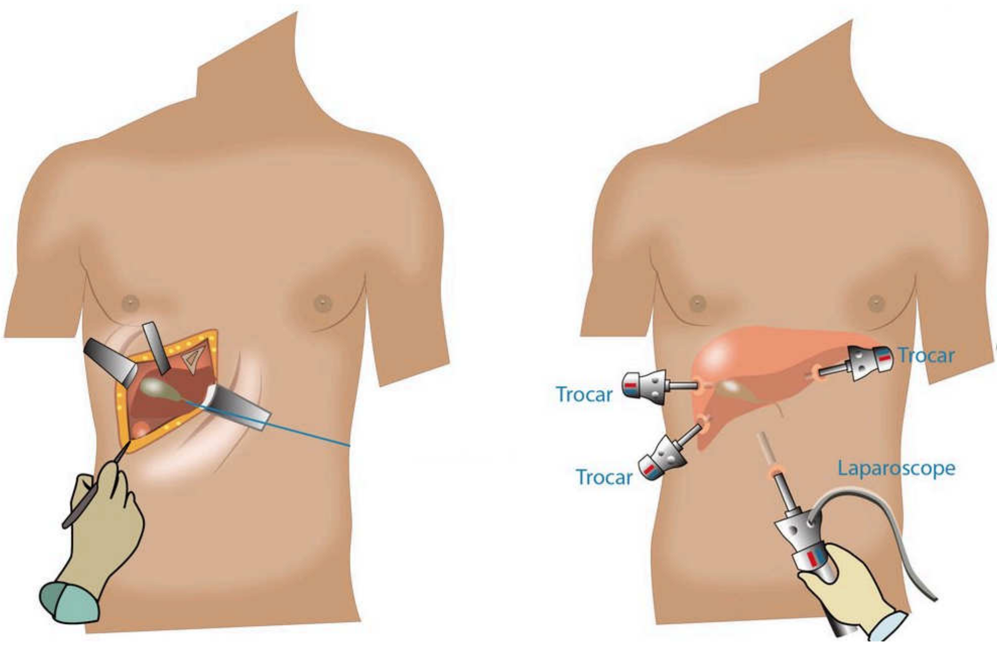


Figure 1.1: Visual comparison of open (left) versus a MIS approach in the context of liver resection (reproduced from [The Liver Center]).

1.1.2 Laparoscopic Liver Resection (LLR)

As of today, liver cancer continues to be a formidable global health challenge, ranking as the sixth most common cancer and the third leading cause of cancer-related mortality worldwide [International Agency for Research on Cancer 2024]. With approximately 866,000 new cases and 758,000 deaths recorded annually, and projections suggesting a rise to over 1.3 million cases by 2040 [Rumgay *et al.* 2022], the demand for effective intervention has never been higher. To combat this issue, surgical resection remains the cornerstone of curative-intent treatment. Within this context, LLR has transitioned into a primary treatment modality [Wakabayashi *et al.* 2015].

LLR represents one of the most technically demanding frontiers within the field of minimally invasive hepatobiliary surgery [Coelho 2016, Jia *et al.* 2018]. While the first laparoscopic liver procedures were reported in the early 1990s [Morise & Wakabayashi 2017, Reich *et al.* 1991], the adoption of LLR was initially slower than that of laparoscopic cholecystectomy or colectomy [Buell *et al.* 2009, Hallet *et al.* 2016] due to the liver's complex vascular architecture and the inherent risk of uncontrollable hemorrhage [Hallet *et al.* 2016, Jia *et al.* 2018].

However, over the last two decades, advancements in surgical instrumentation, such as ultrasonic dissectors and vascular staplers, combined with improved anesthetic management, have transitioned LLR from an experimental approach to a routine practice for specific indications, such as lesions in the peripheral segments [Wakabayashi *et al.* 2015, Fretland *et al.* 2018]. The shift toward LLR is driven by significant improvements in patient recovery compared to traditional OLR. These clinical advantages include smaller incisions, which lead to decreased postoperative pain and lower analgesic requirements, shorter hospital stays, and a quicker return to functional independence, lower rates of incisional hernias and wound infections, and notably, a reduction in postoperative ascites for patients with underlying cirrhosis [Fretland *et al.* 2018, Morise *et al.* 2015].

Despite its benefits, LLR introduces a set of constraints that complicate the surgeon's task. Unlike OLR, where a surgeon can manually palpate the liver to locate deep-seated tumors or identify the course of major vessels, LLR results in a complete loss of haptic feedback [Alleblas *et al.* 2017]. The liver is a voluminous, non-rigid organ characterized by a highly complex internal network of hepatic veins, portal branches, and biliary ducts [Strasberg 2005, Abdel-Misih & Bloomston 2010]. In a laparoscopic setting, the surgeon's view is restricted to the surface of the organ, while the targets for resection and the vessels to avoid are buried deep within the organ [Hanifati *et al.* 2026].

Currently, the success of LLR depends heavily on the surgeon's ability to perform mental registration, which is the cognitive process of mapping preoperative 3D images onto the 2D laparoscopic view on the monitor [Ramalhinho *et al.* 2023]. This process is inherently flawed due to:

- Organ deformation: The liver deforms significantly due to pneumoperitoneum (insufflation gas), patient positioning, and surgical manipulation [Nainamalai *et al.* 2025].
- Visual limitations: The 2D camera lacks depth perception, making it difficult to gauge the exact distance to a hidden tumor [Dirie *et al.* 2018].

1.1.3 The IMMORTALLS Project

This thesis is conducted within the framework of the Intraoperative MultiModal Real-time Tumour Augmentations for Laparoscopic Liver Surgery (IMMORTALLS) project from Pascal Institute, involving a multidisciplinary collaboration with clinical partners (Centre Hospitalier Universitaire (CHU) Clermont-Ferrand and CHU Saint-Etienne), research laboratories (Institut Pascal, SIGMA-Clermont), and industrial expertise (SURGAR). The principal investigator of the project is Erol Ozgur, and it is funded by ANR JCJC. This project aims to overcome the fundamental barriers preventing the widespread adoption of Augmented Reality (AR) in liver surgery.

The primary objective of IMMORTALLS is to enable automatic and real-time AR guidance during LLR. While previous initiatives, such as the Hepataug project (2015–2018) [Espinel *et al.* 2020, Le Roy *et al.* 2019], successfully demonstrated the feasibility of registering preoperative Computed Tomography (CT) volumes to laparoscopic views, they were often limited by a lack of automation and by an inability to account for real-time intraoperative liver deformations.

The IMMORTALLS project addresses these limitations by shifting from a purely *predictive* model, based solely on preoperative data, to a *measured* model that incorporates intraoperative feedback. The core hypothesis is that Laparoscopic Ultrasound (LUS) can serve as the missing link, providing high-frequency, local measurements of the tumor's actual position despite the organ's deformation or partial visibility.

The project is structured around two major technical pillars:

- **Work Package 1:** Multimodal Registrations and Augmentations
  This package focuses on the mathematical and computational challenge of aligning three distinct coordinate frames: the preoperative 3D CT volume, the 2D intraoperative LUS slice, and the 2D monocular laparoscopic video. It specifically targets the partial visibility problem, where often less than 30% of the liver is visible to the camera, by using the LUS probe as a local reference to anchor the global CT model. This thesis is within this Work Package.
- **Work Package 2:** Robotic LUS Visual Servoing
  Recognizing that manually manipulating a LUS probe while operating is cognitively demanding, the project proposes a robotic control system. This system uses multimodal sensing (combined data from the laparoscope and the LUS image) to automatically track the tumor and maintain an optimal imaging plane, thereby reducing the surgeon's workload.

1.2 Intraoperative Navigation

In this section, the principal data modalities used in MIS are presented. The discussion begins with intraoperative imaging and sensing modalities and subsequently proceeds toward preoperative modalities.

1.2.1 Laparoscope

A laparoscope is a rigid endoscopic imaging system that serves as the primary visual sensor in MIS, enabling real-time visualization of internal anatomy through small incisions. Modern laparoscopes typically employ rod-lens optical assemblies coupled with high-definition digital cameras and fiber-optic illumination to provide high-quality intraoperative video streams [Maier-Hein *et al.* 2017]. Despite advances in hardware, laparoscopic imaging is inherently affected by a constrained field of view, non-uniform illumination, specular highlights, and limited depth perception in monocular systems, all of which introduce significant challenges for image interpretation and analysis [Wang *et al.* 2022]. These limitations directly impact downstream medical image processing tasks such as 3D reconstruction, surgical scene segmentation, and augmented reality registration, where robustness to visual artifacts and geometric distortions is essential [Wang *et al.* 2022]. Consequently, recent research emphasizes the importance of explicitly modeling the laparoscope's geometric and photometric properties [Makki *et al.* 2023, Chandelon & Bartoli 2024], as well as leveraging temporal and spatial priors, to improve accuracy and reliability in computer-assisted surgical applications [Rabbani *et al.* 2022b].

1.2.2 Laparoscopic Ultrasound (LUS)

LUS is an intraoperative imaging modality that integrates a miniaturized Ultrasound (US) transducer with a laparoscopic probe, enabling real-time subsurface visualization of anatomical structures during minimally invasive surgery. Unlike a standard laparoscope, which is limited to surface imaging, LUS provides access to internal tissue characteristics such as vascularization, lesion boundaries, and organ parenchyma, thereby enhancing surgical decision-making and intraoperative guidance [Lubner *et al.* 2021, Santambrogio *et al.* 2022]. However, LUS imaging is inherently affected by modality-specific challenges, including speckle noise, acoustic shadowing, limited field of view, and strong operator dependency, which complicate interpretation and downstream processing [Noble & Boukerroui 2006].

Fig. 1.2 demonstrates an example of US integration in Operating Room (OR).

1.2.3 Computed Tomography (CT)

CT is an imaging modality based on the use of X-rays, a form of electromagnetic ionising radiation, to visualize internal anatomical structures. As X-rays traverse the body, they undergo attenuation that varies according to the physical properties

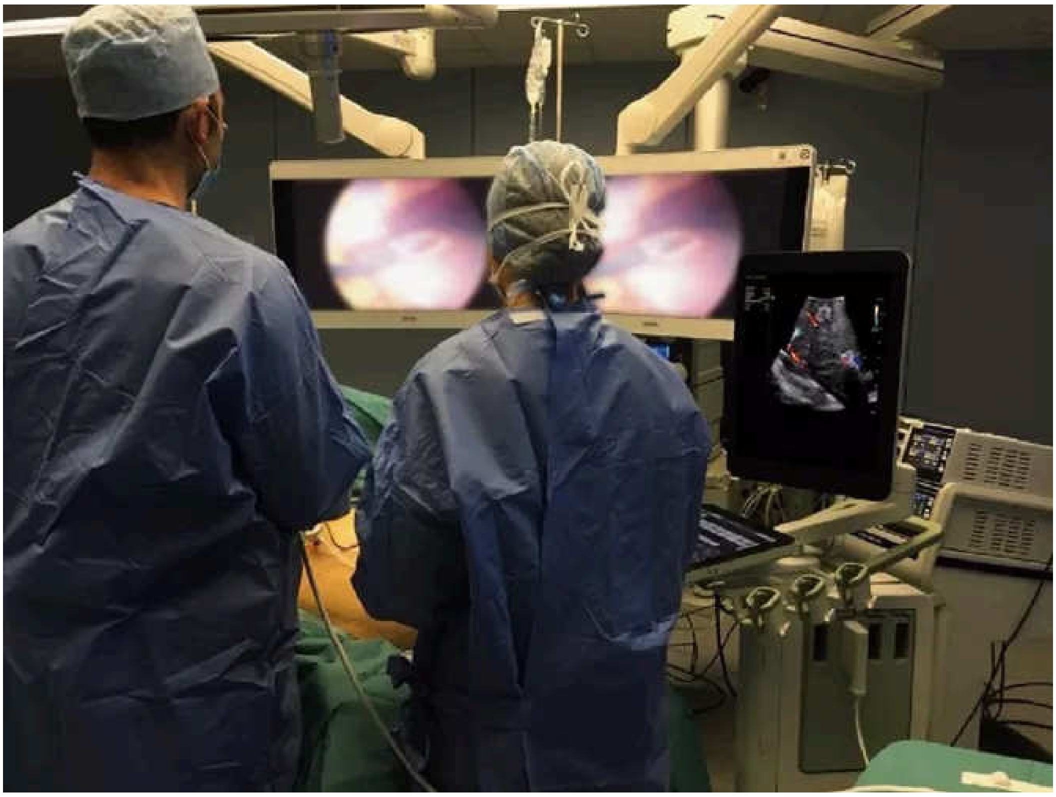


Figure 1.2: Standard clinical workflow for intraoperative ultrasound (US) deployment, featuring an independent machine and display setup. Source: [GE Healthcare 2023].

of the tissues encountered, such as density and atomic composition. This differential absorption enables discrimination between structures such as bone, soft tissue parenchyma, and pathological formations, including tumors.

A CT scanner is composed of a circular gantry that houses both the X-ray source and a corresponding detector array, along with a motorized patient table. During image acquisition, the patient lies in a supine position while the X-ray tube emits a collimated fan beam and rotates continuously around the body. The transmitted radiation, attenuated by the tissues, is captured by detectors positioned opposite the source and subsequently processed by a reconstruction algorithm to generate cross-sectional images [Jung 2021, Labrunie 2024].

Each reconstructed image corresponds to a tomographic slice with a finite thickness, typically ranging from 1 mm to 10 mm [Abdulkareem *et al.* 2023]. These slices are composed of volumetric pixels, or voxels, whose intensity values are proportional to the degree of X-ray attenuation within the corresponding tissue volume. By incrementally advancing the patient table, successive slices are acquired and combined to form a 3D volumetric dataset, commonly used in preoperative planning and diagnostic assessment [Jung 2021, Labrunie 2024].

1.2.4 Magnetic Resonance Imaging (MRI)

Magnetic Resonance Imaging (MRI) is a non-invasive diagnostic modality that leverages the principles of nuclear magnetic resonance to produce high-resolution anatomical and functional images. The fundamental physical process involves the interaction between a strong external magnetic field, typically denoted as B_0, and the intrinsic magnetic moments (spins) of hydrogen nuclei (^{1}H) within the body. When an Radiofrequency (RF) pulse is applied at the Larmor frequency, these spins are excited from their equilibrium state; upon removal of the pulse, they undergo relaxation processes characterized by T_1 (longitudinal) and T_2 (transverse) time constants. As noted in [Lauterbur 1973], the introduction of spatial magnetic field gradients allows for the encoding of position into the signal frequency and phase, enabling the reconstruction of 2D and 3D images through Fourier transformation.

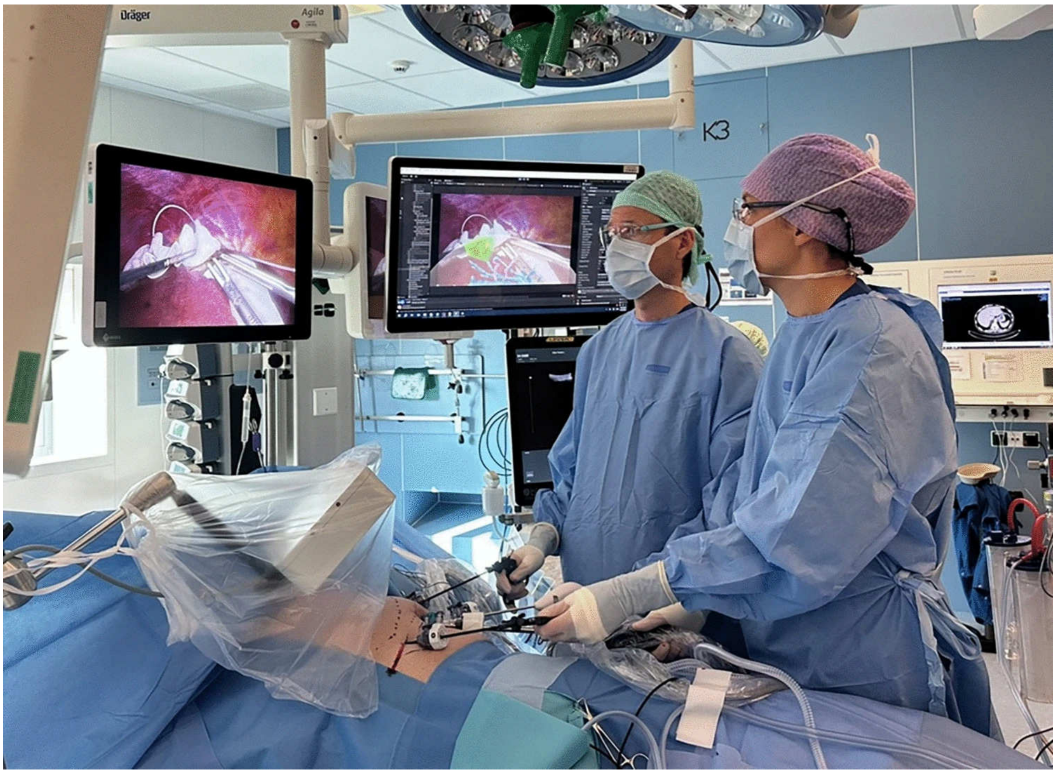


Figure 1.3: As exemplified by the OR configuration retrieved from [Olthof *et al.* 2026], intraoperative data stream visualization traditionally relies on distributed independent monitors. In this specific configuration, the left monitor renders the standard laparoscopic view, whereas the right monitor visualizes the integrated AR guidance. Preoperative images remain accessible on the background display to the far right.

1.3 Related Works on Surgical Guidance

Surgical guidance has evolved significantly over the past decades, driven by the need to enhance intraoperative perception, precision, and decision-making. This section reviews the state-of-the-art in surgical guidance systems, progressing from general surgery to MIS, and finally to LLR, with a particular emphasis on AR-based approaches. A more detailed, chapter-specific literature review will be provided in subsequent sections of this thesis.

1.3.1 General Surgery

Image-Guided Surgery (IGS) has become a cornerstone in modern operative practice, particularly in neurosurgery, orthopedics, and oncologic surgery [Lin *et al.* 2023, Wang *et al.* 2015]. These systems typically integrate preoperative imaging modalities such as CT or MRI with intraoperative tracking systems to provide spatial localization of instruments relative to patient anatomy.

Early navigation systems relied heavily on rigid registration and optical tracking, enabling accurate localization but often requiring surgeons to shift attention between the operative field and external displays. This limitation introduced cognitive load and disrupted surgical workflow. Consequently, research has increasingly focused on improving intraoperative visualization and reducing attention shifts [Lin *et al.* 2023].

AR has emerged as a natural evolution of IGS, aiming to overlay virtual anatomical models directly onto the surgical field. A systematic review of optical see-through head-mounted displays highlights that surgical guidance is the most common AR application, with strong emphasis on preoperative model visualization and intraoperative alignment challenges [Birlo *et al.* 2022]. These systems attempt to address the *mental mapping* problem inherent in conventional navigation, where surgeons must mentally reconstruct 3D anatomy from 2D images.

Despite promising developments, key limitations persist in general surgical guidance systems, including registration accuracy and robustness, handling of soft tissue deformation, and ergonomic and perceptual challenges (e.g., depth perception, occlusion). These challenges become even more pronounced in minimally invasive contexts.

1.3.2 Minimally Invasive Surgery

Although MIS has demonstrated clear clinical benefits, these advantages come at the cost of increased technical difficulty and reliance on visual cues. MIS, particularly laparoscopy, introduces additional constraints that complicate surgical guidance. These include a restricted field of view, absence of haptic feedback, and limited depth perception. As a result, MIS has been a major driver for innovation in image-guided and AR-based systems.

Traditional guidance in MIS relies on:

- LUS for subsurface visualization,
- Preoperative imaging displayed on external monitors,
- Surgeon experience for spatial reasoning.

These approaches require cognitive integration of multiple information sources, which is inherently inefficient. To address this, AR-based systems aim to fuse intraoperative video with preoperative models or real-time imaging.

For example, AR systems integrating US with laparoscopic video have demonstrated the feasibility of real-time image fusion, allowing surgeons to visualize subsurface structures directly within the operative field [Lau *et al.* 2019]. Similarly, intraoperative CT-based navigation systems have demonstrated the potential of guidance in MIS [Peek *et al.* 2024].

Fig. 1.3 displays an example configuration of the various imaging modalities and an AR system incorporated into the OR workflow.

1.3.3 Laparoscopic Liver Resection

LLR represents one of the most challenging applications of MIS due to the liver's complex vascular structure, deformability, and the lack of tactile feedback. Although LLR offers significant benefits, such as reduced morbidity and faster recovery, it remains technically demanding and has seen slower adoption compared to other MIS procedures [Schneider *et al.* 2021].

Image-guided navigation systems have been proposed to address these challenges by providing real-time anatomical guidance. The several categories of navigation systems include [Schneider *et al.* 2021]:

- US-based guidance
- CT or MRI based navigation
- Hybrid systems combining multiple modalities

The review [Schneider *et al.* 2021] emphasizes that while these systems improve anatomical understanding, their clinical adoption is limited by the lack of standardized accuracy metrics, the difficulty in compensating for liver deformation, and workflow integration challenges.

Nevertheless, image-guided navigation constitutes a critical foundation for advanced guidance systems, providing the spatial framework upon which more intuitive visualization techniques, such as AR, can be built.

Building upon image-guided navigation, AR has emerged as a promising approach to improve the intraoperative visualization of anatomical information in LLR. By directly overlaying preoperative 3D models [Mhiri *et al.* 2024], or US image [Rabbani *et al.* 2022a], onto the laparoscopic video stream, AR enables surgeons to access spatially registered information within their natural field of view, thereby reducing the need to mentally correlate separate imaging modalities.

Importantly, AR should be viewed as a complementary extension of image-guided navigation rather than a replacement for it. While AR does not directly address underlying challenges such as registration accuracy or soft tissue deformation, it provides a more intuitive and ergonomic means of presenting navigation data. When coupled with accurate registration and robust tracking, AR has been shown to improve task performance, reduce attention shifts, and facilitate more efficient decision-making during surgery.

As a result, current research increasingly focuses on the integration of AR with advanced registration and modeling techniques, aiming to combine accurate spatial alignment with effective visualization. In this perspective, AR represents a key component in the evolution of surgical guidance systems, contributing to more seamless and information-rich intraoperative environments.

1.4 Objectives, Novelties, and Contributions

This thesis aims to advance intraoperative guidance in LLR by enabling the systematic and seamless integration of LUS into the surgical workflow. The overarching objective is to enhance the surgeon's perception of subsurface anatomical structures through robust AR, thereby addressing the inherent limitations of laparoscopic vision.

To achieve this, the thesis introduces several key methodological contributions. First, a novel algebraic formulation is proposed for the estimation of the 5-Degrees of Freedom (DoF) pose of spherocylindrical LUS probes. This formulation is specifically designed to exploit the geometric constraints of the probe while remaining computationally efficient. Building upon this, a complete real-time processing pipeline is developed and implemented, ensuring the practical applicability of the method in intraoperative settings with strict latency requirements.

Second, the work addresses the inherently underconstrained nature of monocular pose estimation of spherocylinders by proposing two complementary strategies for recovering the missing sixth-DoF. These approaches leverage motion-based constraints and temporal filtering techniques to resolve ambiguities and stabilize pose estimation over time. The resulting framework enables accurate and robust tracking of the ultrasound probe, which is critical for consistent spatial alignment between modalities. This, in turn, supports the first real-time sensor-free augmented reality in the literature by directly fusing intraoperative ultrasound data with the laparoscopic video stream.

Finally, the estimated probe pose and ultrasound-derived information are incorporated as key constraints within a multimodal registration framework. This framework aligns the preoperative model, derived from CT or MRI data, with intraoperative ultrasound and laparoscopic imagery, enabling coherent spatial correspondence across modalities. By doing so, the system facilitates precise full-tumor augmentation, allowing surgeons to visualize critical subsurface structures in situ and in real time.

The proposed methods are extensively validated through a comprehensive and multi-tiered evaluation protocol, encompassing controlled phantom experiments, ex vivo studies, and in vivo clinical surgical datasets. This structured validation strategy enables a progressive assessment of the system under increasing levels of anatomical and procedural complexity, thereby rigorously characterizing its accuracy, robustness, and real-time performance in conditions representative of clinical practice. Beyond methodological advancements, this thesis also contributes to the field through the curation and systematic use of a diverse collection of experimental and clinical datasets, specifically acquired across phantom, ex vivo, and intraoperative scenarios. These datasets are designed to capture the variability inherent to laparoscopic liver surgery, including probe motion dynamics and tissue deformation, and therefore provide a valuable resource for benchmarking and comparative evaluation of multimodal registration and augmented reality approaches. Collectively, the combination of rigorous experimental validation and the introduction of representative datasets not only substantiates the effectiveness and practical feasibility of the proposed framework but also establishes a reproducible foundation for future research, supporting continued progress toward reliable, high-precision intraoperative guidance systems for minimally invasive liver surgery.

1.5 Structure of the Manuscript

The remainder of this thesis is organized into five primary chapters, following a progression from the establishment of evaluation frameworks to the development of sophisticated surgical navigation and augmentation solutions.

- **Chapter 2: Evaluation Resources: Datasets and Error Metrics**

  This chapter establishes the foundational framework for the experimental validation used throughout the thesis. It reviews existing datasets and introduces several novel contributed datasets, including clinical, ex vivo, semi-synthetic, and phantom-based data. Furthermore, it defines the specific error metrics used to quantify the accuracy of LUS probe pose estimation and tumor target registration.

- **Chapter 3: Single-shot 5-DoF Laparoscopic Ultrasound Probe Pose Estimation**

  Focusing on the challenge of vision-based tool tracking, this chapter proposes a methodology for estimating the 5-DoF pose of an LUS probe from a single camera frame. It details the solution pipeline for probe segmentation and contour extraction, concluding with an analysis of sensitivity and performance across phantom and patient-data studies.

- **Chapter 4: Laparoscopic Ultrasound Probe Pose Disambiguation, Filtering, and Imaging Plane Augmentation**

  Building upon the previous chapter, this chapter introduces constraints to achieve full 6-DoF pose estimation. By utilizing shaft and motion constraints, the proposed system filters noise and resolves pose ambiguity. This chapter also introduces the integration of AR rendering to visualize the US imaging plane within the laparoscopic view.

- **Chapter 5: Multimodal Registration and Full-tumor Augmentation**

  This chapter presents the culminating technical contribution: a framework for the multimodal registration of preoperative imaging with intraoperative ultrasound. It details the modeling assumptions, cost functions, and minimization methods required to achieve full-tumor augmentation, supported by comprehensive qualitative and quantitative studies.

- **Chapter 6: Conclusion and Future Research**

  The final chapter summarizes the primary contributions of this research and discusses its implications for the field of LLR. It concludes by identifying remaining challenges and suggesting potential avenues for future investigation in surgical navigation systems.

Chapter 2

Evaluation Resources: Datasets and Error Metrics

Contents

2.1 Introduction

The validation of intraoperative navigation and guidance systems is a critical requirement in the medical field, yet it remains a formidable challenge. In fact, within this domain, establishing a robust validation framework might prove more difficult and complex than the development of the methodological algorithms themselves. Establishing accuracy for AR and pose estimation is particularly difficult because direct Ground Truth (GT) is often unavailable or impossible to capture during live surgical procedures. Acquiring a rigorous GT in the OR using real patient data is rarely feasible; it necessitates altering established surgical routines, introducing new tracking devices into a highly constrained environment, overcoming strict sterility issues, and imposing an unacceptable mental burden on the surgical team.

Because of these immense practical hurdles, the design and implementation of a comprehensive validation infrastructure is, *per se*, a considerable scientific contribution. To address this critical gap and ensure the reliability of system enhancements, this chapter introduces meticulously acquired datasets. The conceptualization, acquisition, and development of these datasets required significant time and effort to successfully navigate the complex constraints of the surgical environment. Ultimately, this dataset represents a highly valuable and significant resource for the field, providing a rare and essential foundation for bridging the gap between algorithmic development and clinical translation.

To systematically address these formidable validation hurdles, this thesis utilizes, builds upon, and contributes a comprehensive suite of evaluation resources:

- **Existing Benchmarks:** We leverage high-value clinical LUS-based datasets, such as those from [Rabbani *et al.* 2022a], which use a pose estimation pipeline to recover tumor locations.

- **3D-Printed Phantoms:** Based on the work of [Ribeiro *et al.* 2024], these phantoms provide a patient-derived anatomical baseline with exact geometric GT for both surface and subsurface anatomy under controlled deformations.

- **Contributed Datasets:** This work introduces a multimodal dataset spanning clinical videos, a high-precision ex vivo sheep liver model, and dedicated phantom-based sequences.

- **Ground Truth & Tracking:** Precise GT is achieved through ArUco marker-based tracking and calibrated acquisition setups, enabling 6-DoF pose annotation for the LUS probe.

- **Uncertainty Quantification:** To ensure the reliability of our experimental results, we perform a rigorous quantification of the GT uncertainty. By analyzing the distribution of residuals under mechanical constraints, we establish the intrinsic measurement noise of the system.

- **Scalable Deep Learning (DL) Development:** We developed a semi-synthetic generation framework implemented as PyTorch-compatible dataloaders. These pipelines allow for the creation of an unbounded volume of training samples for tasks like LUS probe segmentation and pose estimation.

By synthesizing these varied evaluation resources, from complex clinical environments to highly controlled phantoms with definitive uncertainty metrics, this chapter delivers the essential validation infrastructure for this thesis. Ultimately, this rigorous framework provides the reliable baseline required to thoroughly evaluate and justify the AR enhancements and registration techniques introduced in the following chapters.

2.2 Existing Datasets

Among the existing research efforts in LLR and the developed datasets, two preceding works from the group were particularly valuable for this thesis, notably by providing datasets that significantly facilitated the experimental design and validation protocols.

2.2.1 Clinical LUS-Based Dataset for Registration Evaluation

One of the most valuable existing datasets for LLR is [Rabbani *et al.* 2022a] from the preceding laboratory's research. This dataset is acquired in real clinical conditions during LLR and is specifically designed to enable quantitative evaluation of registration accuracy using intraoperative ground truth.

The central idea of the data acquisition process is to use LUS as a means to recover the spatial location of tumors, which are otherwise not directly visible in laparoscopic images. To achieve this, a complete calibration and tracking pipeline is established to express LUS observations within the laparoscopic coordinate system.

The acquisition pipeline consists of three main stages. First, a 3D reconstruction of a checkerboard pattern attached to the LUS probe head is performed using Structure from Motion (SfM) techniques. This provides an accurate geometric reference for the probe. Second, the LUS probe is calibrated using a dedicated phantom containing wire structures, enabling estimation of the transformation between LUS image coordinates and the probe coordinate system, along with pixel scaling factors. Third, the pose of the LUS probe relative to the laparoscopic camera is estimated using the observed checkerboard, allowing LUS data to be mapped into the laparoscopic reference frame.

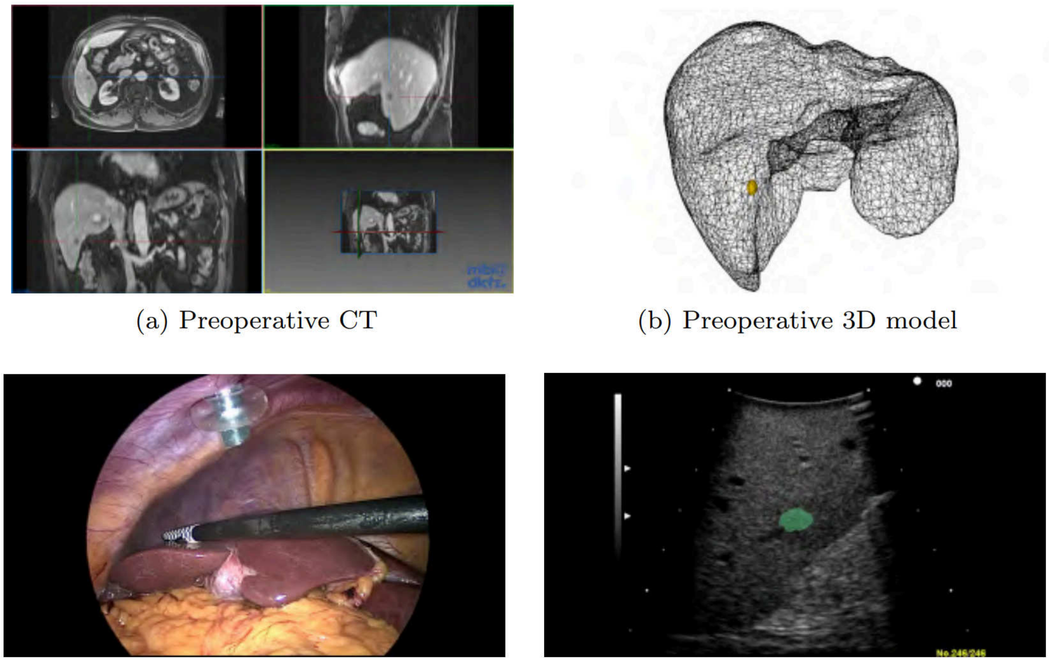
This pipeline enables the transfer of tumor contours identified in LUS images into a common 3D coordinate system aligned with the laparoscopic view, effectively providing GT spatial information.

The resulting clinical dataset includes four patients, each presenting a single liver tumor. For each case, the dataset contains:

- A preoperative CT scan,
- A preoperative 3D model of the liver and tumor obtained via manual segmentation by an expert surgeon,
- A set of intraoperative laparoscopic images,
- Corresponding LUS images acquired synchronously with the laparoscopic data.

In the ultrasound images, tumor boundaries are manually delineated and provided as binary masks. Additionally, the dataset includes all necessary calibration and pose estimation data, enabling full reproducibility of the GT computation and evaluation pipeline.

Fig. 2.1 outlines the available modalities for each patient within the dataset. Crucially, each instance also includes the spatial pose data for the LUS plane.



(a) Preoperative CT (b) Preoperative 3D model

(c) Laparoscopic images (d) Segmented LUS images

Figure 2.1: Data modalities available per patient in the dataset provided by [Rabbani *et al.* 2022a]. Diagram retrieved from the original study.

2.2.2 3D-Printed Phantoms for Registration Evaluation

To create the Phantom datasets for this thesis, the 3D-printed phantoms of [Ribeiro *et al.* 2024] are used. These phantoms are derived from a real patient-specific anatomical model, ensuring that the baseline geometry reflects clinically realistic liver morphology.

Starting from this original mesh, a set of eight distinct deformed configurations is generated to simulate plausible intraoperative shape variations. These deformations allow the creation of multiple instances of the same anatomical structure under controlled geometric transformations. Each configuration is then physically fabricated using additive manufacturing, resulting in a collection of phantoms with known correspondences and controlled variability.

The phantoms incorporate both external surface features and internal structures, such as a set of virtual tumor inclusions, whose geometry and spatial location are fully defined during the design process. This ensures the availability of exact GT for both surface and subsurface anatomy.

In addition, a set of fiducial landmarks is embedded or defined on the phantoms. These landmarks serve a dual purpose. First, they enable accurate rigid registration between different acquisitions or between the phantom and its reference model. Second, they provide a means to quantitatively validate deformable registration by allowing point-wise comparison between corresponding anatomical locations across different deformed states. The sequential methodology employed to fabricate each phantom is illustrated in Fig. 2.2.

Data acquisition is performed using imaging modalities consistent with the clinical setup, such as optical imaging and an ultrasound probe. This ensures that the dataset remains representative of the sensing conditions encountered in practice while benefiting from the repeatability of a controlled environment.

Although the mechanical properties of the printed materials do not fully replicate the complexity of living tissue, the proposed phantom dataset provides an essential intermediate validation stage. It enables systematic evaluation of registration algorithms under known geometric transformations, isolating algorithmic performance from the uncertainties inherent in clinical data.

Overall, these phantoms offer:

- A patient-derived anatomical baseline,
- Multiple controlled deformation states (eight configurations),
- Exact geometric ground truth,
- Landmark-based validation for both rigid and deformable registration,
- High repeatability and experimental control.

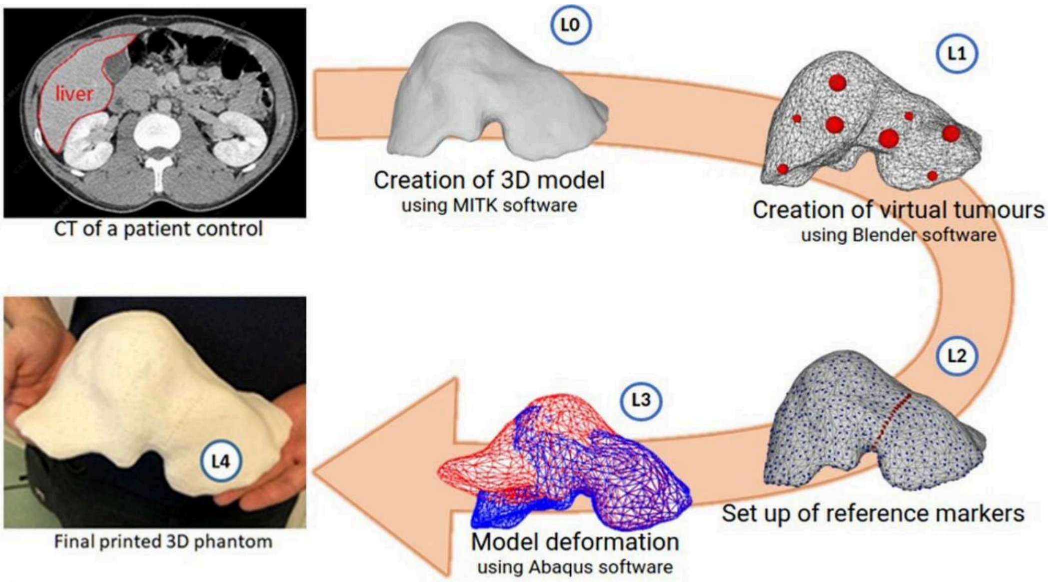


Figure 2.2: Chronological workflow for phantom fabrication as described by [Ribeiro *et al.* 2024]. Diagram adapted from the original study.

2.3 Contributed Datasets

To further validate the proposed methods under diverse experimental conditions, and to address limitations in existing datasets, particularly the absence of timestamps and temporally consecutive frame sequences, several additional heterogeneous datasets were created, including clinical, ex vivo, semi-synthetic, and phantom datasets. These datasets are introduced and described in detail in this section, and summarized in Table 2.1.

2.3.1 Clinical

We begin with the clinical dataset, which integrates different modalities acquired from real patients undergoing LLR.

2.3.1.1 Videos

As part of this work, a dedicated dataset of LLR procedures was constructed through both prospective acquisition and retrospective curation.

First, a prospective data collection was conducted during five LLR surgeries. For each procedure, multiple synchronized data streams were recorded. These included the primary laparoscopic video, and in two robot-assisted cases, dual video channels were captured simultaneously. In addition, a checkerboard pattern was recorded intraoperatively to enable accurate camera calibration. US imaging was also acquired in a synchronized manner using the LUS probe. Complementary to the intraoperative data, preoperative imaging data were collected, along with corresponding 3D reconstructions of the liver anatomy, including the organ surface, tumor, and vena cava.

Second, a retrospective exploration of the hospital's surgical video archive was performed to expand the dataset. The selection criterion focused on the presence of the LUS probe within the recorded procedures. Through this process, a total of 19 additional LLR surgeries were identified and included.

Together, these efforts resulted in a multimodal dataset combining intraoperative video, ultrasound imaging, and preoperative anatomical models, providing a comprehensive foundation for subsequent analysis and method development.

2.3.1.2 Images

From the collected dataset, a subset of 12 surgeries was selected for further processing. For these procedures, laparoscopic video streams were systematically sampled to extract individual frames. In collaboration with SURGAR, a curation process was carried out to identify distinctive and informative frames within each surgery.

This process resulted in a total of 3,144 images annotated with segmentation masks of the LUS probe. In addition, more than 40,000 frames not containing the LUS probe were collected. These background images were subsequently utilized in the generation of a semi-synthetic dataset.

Fig. 2.3 demonstrates an exemplary instance of the synchronized data streams available within the dataset.

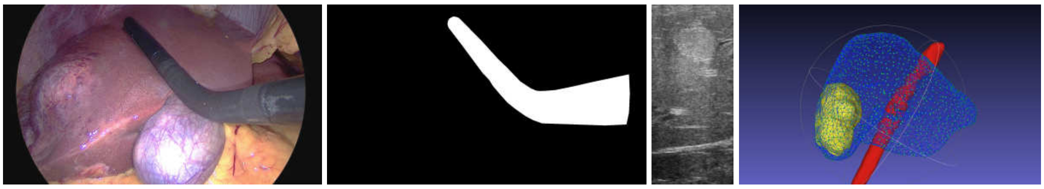


Figure 2.3: Representative keyframe extracted from the surgical video dataset. From left to right, the panels display the intraoperative surgical image, the LUS segmentation, the synchronized ultrasound stream, and the preoperative anatomical model highlighting the tumor (yellow) and vena cava (red). The recorded sequence for this specific procedure also includes camera calibration data and baseline background images acquired without the LUS probe.

2.3.2 Ex Vivo

A third dataset was acquired in a controlled ex vivo setting using a sheep liver, chosen for its anatomical resemblance to the human liver [Chanwat *et al.* 2025]. This dataset consists of two laparoscopic video sequences designed to provide accurate GT for methodological validation.

In this setup, the LUS probe pose was tracked in 6-DoF using an ArUco marker-based tracking system [Garrido-Jurado *et al.* 2014] combined with a calibrated acquisition setup. This enabled precise spatial GT annotation of the probe throughout the sequences. In addition, pixel-wise segmentation annotations were generated for both the LUS probe head and shaft, separately. Camera calibration parameters were also estimated and provided as part of the dataset to ensure geometric consistency across experiments. Fig. 2.4 demonstrates a frame of the dataset along with the GT marker extrinsic calibration approach.

2.3.3 Semi-synthetic

Next, a semi-synthetic dataset generation framework was developed to augment the available real data. Two complementary data generation pipelines were designed, both implemented as PyTorch-compatible dataloaders, allowing seamless integration into standard deep learning training workflows. Owing to their stochastic design, these pipelines can theoretically generate an unbounded number of training samples.

2.3.3.1 Segmentation Pipeline

The first pipeline targets LUS probe segmentation. It takes as input a set of original images with corresponding LUS segmentation masks, along with a separate collection of background images. During data generation, an LUS instance is randomly

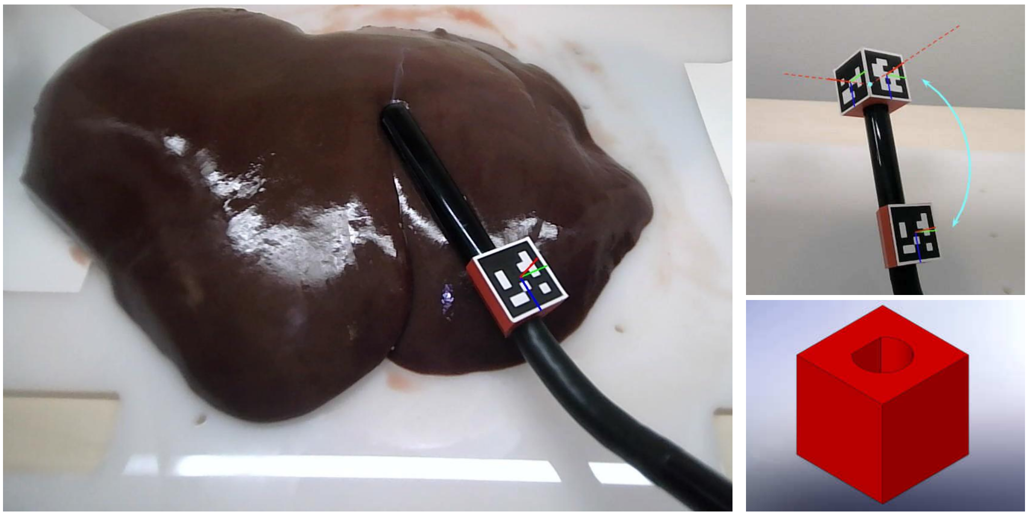


Figure 2.4: Left: a cropped frame from the ex vivo dataset, along with the added coordinate frame on the GT ArUco marker for verification. Top-right: calibration of the attached ArUco marker with the LUS body through the 3D-printed calibration cube, whose center is precisely aligned with the LUS tip's center, and the upper surface is parallel to the sensor. Bottom-right: CAD rendering of the designed calibration cube.

selected, cropped using its ground truth mask, and composited onto a randomly chosen background image at a random location. This process produces realistic surgical scenes containing a synthetically placed LUS probe, along with the corresponding ground truth segmentation mask. The resulting samples can be directly used to train segmentation networks in a supervised manner.

2.3.3.2 Pose Estimation Pipeline

The second pipeline is designed for DL-based LUS pose estimation. It operates by first sampling a background image, onto which a virtual LUS probe is rendered. Given the probe head pose, camera intrinsic parameters, and scene lighting conditions, a realistic rendering of the LUS head is generated and composited into the image. This approach provides precise control over the 5-DoF pose GT. In addition to pose annotations, the rendering process also produces the corresponding LUS silhouette, enabling its use for both pose estimation and segmentation tasks.

To mitigate the domain gap between synthetic and real data, several strategies were incorporated into the generation process. In the segmentation pipeline, alpha blending and boundary smoothing were applied when compositing the LUS probe onto background images to reduce visual discontinuities. Furthermore, photometric augmentations such as variations in brightness, contrast, and color distribution were introduced to better reflect the variability observed in real surgical environments. In the rendering-based pipeline, realism was enhanced by incorporating lighting variations consistent with laparoscopic scenes, as well as accounting for camera intrinsic parameters during projection. Additional perturbations, including noise injection and slight geometric inconsistencies, were also introduced to improve robustness. These design choices aim to narrow the sim-to-real gap and improve the generalization capability of models trained on the generated data.

Although the semi-synthetic data generator operates natively as a dataloader, it additionally supports exporting and saving the generated datasets, as illustrated in Fig. 2.5.

Figure 2.5: Overview of the semi-synthetic data generation pipeline. Utilizing background imagery, LUS pose data, camera parameters, and illumination conditions as inputs (left), the pipeline renders a synthetic LUS probe onto the image (middle). Concurrently, it exports the ground truth (GT) pose and LUS mask (right) to support downstream processing pipelines.

2.3.4 Phantom

Finally, two phantom-based datasets were constructed to enable controlled experimentation under known conditions.

The first dataset was acquired using a deformable silicone liver phantom scaled to approximately half the size of a real human liver, designed to mimic tissue deformation and probe-induced pressure. It consists of two video sequences, for which accurate 6-DoF LUS probe GT pose (see Sec. 2.3.2) and camera calibration parameters are provided.

The second dataset is based on rigid liver phantoms introduced in Sec. 2.2.2. It captures LUS probe motion over these phantoms and similarly provides GT probe pose and camera calibration data. By registering the phantom geometry to its corresponding virtual model, the positions of virtual tumors can be recovered. This also enables the generation of corresponding virtual ultrasound images, which are exploited in selected experiments. Two selected frames from these datasets can be observed in Fig. 2.6.

Figure 2.6: Left: a frame of the dataset with the deformable silicone phantom. Right: A frame from the created dataset with the rigid phantoms of [Ribeiro *et al.* 2024].

Table 2.1: Summary of the Contributed Datasets

Dataset	Modality	GT/Annotation	Technical Features
Clinical	MIS video; US video; Preoperative model	3K+ probe masks; 40K+ bkg. frames	Camera calibration parameters available for some of them.
Ex Vivo	MIS video	6-DoF LUS GT pose; Probe head & shaft segmentation	LUS contact verified by US images; camera calibration.
Semi-synthetic	MIS images	5-DoF LUS GT pose; LUS masks	Alpha blending, photometric augmentation, and lighting control; camera calibration.
Silicone Phantom	MIS video	6-DoF LUS GT pose; Probe head & shaft segmentation	Camera calibration, LUS contact verified by US images.
Printed Phantom	MIS video; US video; Preoperative model	6-DoF LUS GT pose; GT liver deformation	Deformable simulation; virtual tumors; camera calibration.

2.4 Error Metrics

In this section, the different error metrics used throughout the experiments are defined, including those employed for evaluating pose estimation accuracy and registration performance.

2.4.1 Laparoscopic Ultrasound Probe Pose

To define LUS probe pose error, first we model the geometry of the LUS probe as illustrated in Fig. 2.7. The probe head is actuated by two distal levers, denoted *l1* and *l2*, which are not depicted in Fig. 2.7. Lever *l1* controls the up–down articulation of the probe head, while lever *l2* enables left–right rotation relative to the shaft. In Fig. 2.7, $\underline{\mathbf{x}}, \underline{\mathbf{y}}, \underline{\mathbf{z}}$ define the axes of the camera coordinate frame with origin $\mathbf{o}$. The vector $\mathbf{t}$ denotes the probe tip position (i.e., the center of the hemispherical head), and the unit vector $\underline{\mathbf{h}}$ specifies the orientation of the probe head axis. Likewise, the unit vector $\underline{\mathbf{s}}$ represents the orientation of the probe shaft axis. When the lateral articulation of the probe head is held constant, the cross product $\underline{\mathbf{h}} \times \underline{\mathbf{s}}$ defines the normal to the LUS image plane (shown in blue). The vector $\underline{\mathbf{m}}$ denotes the normal of the interpretation plane spanned by the head axis line ℓ_h and the origin $\mathbf{o}$, while $\underline{\mathbf{n}}$ denotes the normal of the interpretation plane spanned by the shaft axis line ℓ_s and $\mathbf{o}$. The rotation angle α, defined about $\underline{\mathbf{m}}$, characterizes the depth variation of the probe head axis direction $\underline{\mathbf{h}}$ within its interpretation plane. Analogously, the rotation angle γ, defined about $\underline{\mathbf{n}}$, characterizes the depth variation of the probe shaft axis direction $\underline{\mathbf{s}}$ within its interpretation plane. The rotation angle θ describes the rotation of the probe head about its own axis $\underline{\mathbf{h}}$. Expressing errors directly in the laparoscope frame is challenging due to their dependence on the probe pose. To address this, we introduce two additional coordinate frames, in addition to the base frame. For position errors, we use the $[\underline{\mathbf{w}}, \underline{\mathbf{t}}, \underline{\mathbf{m}}]$ frame, with the origin at $\mathbf{t}$ where $\underline{\mathbf{t}}$ is the unit vector of $\mathbf{t}$ and $\underline{\mathbf{w}} = \underline{\mathbf{t}} \times \underline{\mathbf{m}}$. The errors along these axes are denoted as *in-plane lateral*, *in-plane depth*, and *out-of-plane*, respectively. For orientation errors, we employ the $[\underline{\mathbf{m}}, \underline{\mathbf{v}}, \underline{\mathbf{h}}]$ frame, defining rotations about each axis as *in-plane*, *out-of-plane*, and *self-rotation*, respectively. By definition, $\underline{\mathbf{v}} = \underline{\mathbf{h}} \times \underline{\mathbf{m}}$.

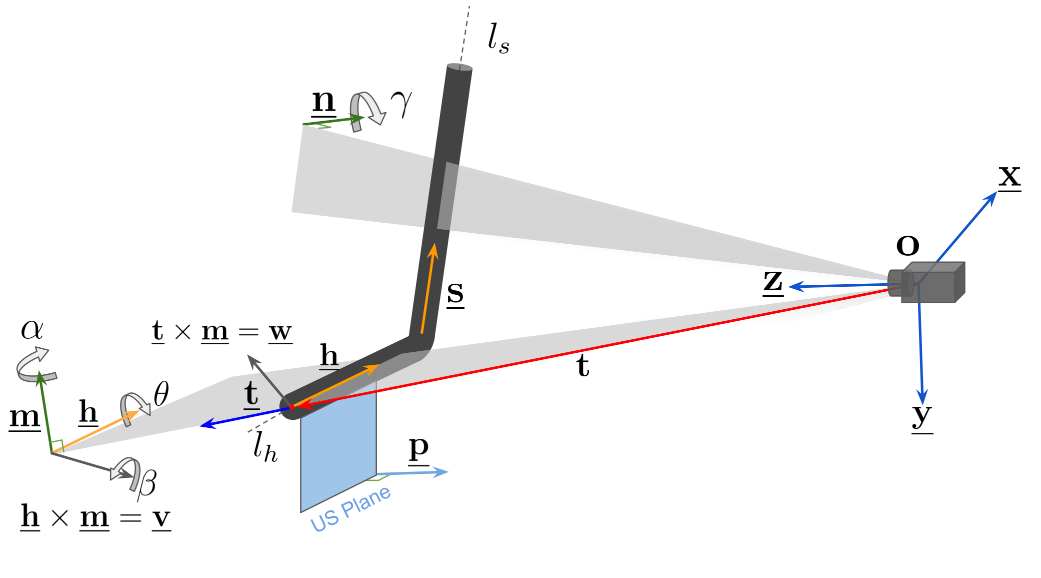


Figure 2.7: Error definition coordinate frames for LUS

2.4.2 Ground Truth Uncertainty Quantification

To assess the accuracy of the ArUco marker-based GT pose estimation in Sec. 2.3.2 and Sec. 2.3.4, an uncertainty quantification experiment was conducted. A linear mechanical constraint was introduced to restrict the motion of the LUS probe, effectively fixing five DoFs while allowing controlled motion along the remaining DoF.

The probe was then moved across multiple poses with respect to the camera, and the corresponding relative transformations were computed. Under ideal conditions, the constrained five DoFs should remain constant (i.e., zero variation). However, deviations from zero were observed due to measurement noise and system inaccuracies.

The distribution of these residuals was analyzed by computing the first and second statistical moments (mean and standard deviation). For the translational components, the mean error was approximately 0 mm with a standard deviation of 1.2 mm. For the rotational components, the mean error was approximately $0°$ with a standard deviation of $0.4°$.

These results provide an estimate of the intrinsic uncertainty associated with the ArUco-based GT, which is subsequently taken into account when interpreting experimental outcomes.

2.4.3 Tumor Target Registration Error

In [Rabbani *et al.* 2022a], to provide a quantitative measure of geometric accuracy, the authors define a variant of the Target Registration Error (TRE) based on the discrepancy between ultrasound-derived tumor contours and the corresponding tumor predicted by registration.

The TRE is computed by sampling points along the tumor contours extracted from LUS and measuring their distances to the registered tumor surface. The formulation involves estimating a deformation field between the two representations, which is decomposed into:

- A rigid component (rotation and translation),
- A residual non-rigid component.

An Iterative Closest Point (ICP) procedure is used to estimate the rigid alignment, after which the residual distances between corresponding points are used to compute the final TRE as an average squared distance over all sampled points.

This formulation allows the TRE to capture discrepancies between:

- Sparse contour observations from ultrasound, and
- Volumetric tumor predictions from the registered preoperative model.

2.5 Conclusion

In conclusion, this chapter has established a comprehensive framework of evaluation resources and error metrics tailored for LLR. By bridging the gap between clinical data and controlled experimental setups, we have created a robust validation pipeline that addresses the inherent difficulties of medical navigation.

The primary outcomes of this chapter include:

- **A Multimodal Data Foundation:** The integration of clinical, ex vivo, and phantom-based datasets ensures that our methods are tested against both real-world complexity and exact geometric GT.
- **Scalable Training Infrastructure:** The introduction of stochastic, semi-synthetic pipelines provides the massive data volume required to train deep learning models for probe segmentation and pose estimation.
- **Reliable Performance Baselines:** Through uncertainty quantification, we have determined that our ArUco-based GT possesses an intrinsic uncertainty of approximately 1.2mm for translation and $0.4°$ for rotation, providing a clear context for interpreting subsequent results.
- **Standardized Metrics:** We have defined specific coordinate frames for measuring LUS probe pose errors and utilized TRE from [Rabbani *et al.* 2022a] to quantify the geometric accuracy of tumor localization.

Collectively, these resources—including over 3,000 annotated clinical images and multiple controlled deformation states—form the essential bedrock for the registration and AR techniques developed in this thesis.

Chapter 3

Single-shot 5-DoF Laparoscopic Ultrasound Probe Pose Estimation

Contents

3.1 Introduction

MIS offers significant advantages over open surgery. It reduces postoperative complications, shortens hospital stays, and lowers overall healthcare costs [Jaffray 2005]. However, MIS prevents direct palpation of organs, making the localization of tumors embedded within soft tissue a major surgical challenge. LUS is a key intraoperative modality used to localize such tumors and is considered the gold standard in liver and kidney surgery. Nevertheless, LUS suffers from several limitations, including a steep learning curve, high operator dependency, a restricted field of view, and inherently low image quality.

Two research directions have recently emerged to mitigate these drawbacks: Robotically-controlled Ultrasound (RobUS) and AR-based guidance [Mohareri *et al.* 2013, Singla *et al.* 2017]. RobUS relieves the surgeon from manually maneuvering the LUS probe to localize and track the tumor, while AR-based guidance fuses LUS images directly into the laparoscopic view. Both approaches are highly promising for surgical assistance; however, they critically depend on quickly and accurately solving the **LUS pose problem**, that is, estimating the position and orientation of the LUS probe in the laparoscope coordinate frame, or equivalently, determining the rigid transformation between the LUS probe and the laparoscope. In most hand-held LUS probes, the transducer **head** is connected to a **shaft** via an articulated joint, and the desired pose refers specifically to the head.

We model the LUS head as a Straight Homogeneous Spherical Cylinder (SHSC), also known as a **spherocylinder** [Dzubiella *et al.* 2000]. Due to the inherent rotational symmetry of the spherocylinder, the pose is ambiguous by 1 DoF: a rotation around the head's cylindrical axis cannot be recovered from single-shot head-only observations. This modeling choice is motivated by the design of many commercial LUS probes, such as the Fujifilm L44LA, Siemens Acuson P300 LP323, Philips L10-4lap, SonoScape LAP7, and Aloka UST-5418, which feature cylindrical shafts and heads with approximately hemispherical tips.

The LUS pose estimation problem is particularly challenging in real OR settings for four main reasons. First, the laparoscope is typically the only available sensor, providing monocular RGB images. Second, LUS probes are generally textureless, reflective, and often partially occluded, which hinders reliable geometric or photometric feature extraction. Third, marker-based or externally tracked solutions (optical or electromagnetic) would require placing additional elements inside the abdominal cavity, which is impractical due to sterilization, safety, and occlusion risks, and would further clutter the OR. Fourth, cable-driven surgical robots cannot provide sufficiently reliable pose estimates from internal encoders, due to inaccuracies in sensing [Bouget *et al.* 2017] and time-accumulated drift [Allan *et al.* 2015]. Even with accurate kinematics, drop-in probes are grasped differently for each intervention, preventing reproducible calibration.

To be practical in the OR, a LUS pose estimation should satisfy four requirements: it must be *(i)* marker-free, *(ii)* sensor-free (beyond the laparoscope), *(iii)* (re)initialization-free, and *(iv)* real-time.

In this chapter, we present the first fully vision-based method for both LUS pose estimation that satisfies all four of these requirements. We name the LUS pose estimation method `Laparoscopic Ultrasound Pose (LUP)`.

The contributions of this chapter are:

- A novel minimal formulation for reconstructing a spherocylinder from its occluding contour,
- A complete algorithmic pipeline, including segmentation and robust model estimation, to recover a 5-DoF head pose.

3.2 Related Work

As this chapter encompasses both geometric foundations and their medical applications, the section is divided into two parts in order to address each aspect separately and comprehensively. Importantly, the concisely focused review of the medical applications directly reflects a pronounced scarcity of existing literature on vision-based LUS pose estimation, highlighting the underexplored nature of this specific research domain.

3.2.1 Vision-based Surgical Tool Pose Estimation

The task of tracking and estimating the pose of surgical instruments, particularly for the da Vinci robotic system, has been approached through several distinct methodologies. The salient ones are grouped into the ensuing categories.

3.2.1.1 Deep Learning-based Methods

The current state-of-the-art has shifted toward markerless, keypoint-based deep learning [Park *et al.* 2024, Xu *et al.* 2025]. These frameworks typically employ Convolutional Neural Network (CNN)s or Transformers to detect discrete tool features, such as tips, joints, or edges, and then map them to 3D space using Perspective n Point (PnP) solvers. While these architectures achieve high precision, they are *data-hungry*, requiring exhaustive 3D-annotated datasets that are difficult to obtain for LUS. Furthermore, the computational overhead of mapping 2D keypoints into 6-DoF poses can introduce significant latency. Many of these deep pipelines struggle to maintain stable performance on standard medical-grade hardware, often failing to reach the frame rates required for seamless surgical integration.

3.2.1.2 Rendering-based Methods

Modern techniques utilize differentiable rendering pipelines to enable end-to-end pose refinement through backpropagation, typically employed to supervise model training [Lu *et al.* 2023]. More recent research, such as [Liang *et al.* 2025], extends this paradigm by incorporating rendering-based optimization into the inference phase to achieve robust intraoperative pose recovery. Such methods are gaining increasing prominence in the literature, driven by recent advances in algorithmic optimization that were not yet available at the time this research was conducted.

3.2.1.3 Model-based Pose Estimation

Model-based approaches utilize the known geometry of the tool to infer its pose. For general robotic tools, [Ye *et al.* 2016] combines image observations with robot joint values to achieve fast run-time performance, while [Allan *et al.* 2015, Allan *et al.* 2018] use optical flow to obtain 3D tracking. However, these are often tailored to specific articulated instruments and are not applicable to LUS probes. For instruments closer in geometry to a LUS probe, [Doignon & Mathelin 2007] proposes a markerless method limited to cylindrical geometries that fails to estimate tip position. Similarly, [Zenteno *et al.* 2021] estimates the pose of a cylindrical fiberscope by maximizing similarity between binary masks [Zenteno *et al.* 2019]. Although this approach is in principle applicable to LUS, it is overly sensitive to segmentation quality and mask noise, often leading to jittery pose estimates.

3.2.1.4 In 2D

Early research focused primarily on localizing tools within the 2D image plane. Tracking surgical tools of the da Vinci robot in 2D images is a well-studied problem using various segmentation and detection frameworks [Sarikaya *et al.* 2017, Laina *et al.* 2017, Kurmann *et al.* 2017]. More recently, enhanced feature-fusion networks have been proposed to improve Frames Per Second (FPS) in complex environments [Liu *et al.* 2022]. While these methods provide essential bounding box or mask information, they do not recover the 3D orientation or depth required for advanced spatial guidance.

3.2.1.5 Other Approaches

A different line of research focuses on registering LUS images to preoperative 3D data, such as CT models [Montaña-Brown *et al.* 2022, Montaña-Brown *et al.* 2021, Ramalhinho *et al.* 2021, Ramalhinho *et al.* 2019]. By aligning the intraoperative LUS feed with the preoperative volume, these methods provide an indirect estimate of the LUS probe pose. While valuable for AR guidance, these registration-based methods are computationally demanding and highly sensitive to initial alignment. Their reliance on complex optimization loops makes them unlikely to operate in real-time under the strict constraints of the OR.

3.2.1.6 The Proposed LUP Method

In contrast to some of the resource-intensive frameworks previously discussed, the proposed `LUP` method is designed to navigate both the clinical and environmental constraints of modern surgical research. Large-scale deep learning models often incur prohibitive training costs, not only in terms of the manual labor required for 3D data preparation but also through their substantial energy consumption. In strict alignment with the lab's environmental ethical code to minimize the carbon footprint of computational research, our approach avoids the continuous retraining of massive, energy-dense architectures.

Furthermore, the inference cost of many state-of-the-art models necessitates high-end GPU clusters. Such hardware is not only power-intensive but also seldom available in standard ORs, where space and power infrastructure are prioritized for life-critical systems. To address these challenges, the `LUP` method adopts a hybrid strategy: it minimizes the use of deep learning to only the essential task of segmentation, subsequently leveraging a rigorous geometrical approach for 6-DoF pose recovery. This design significantly lowers the computational profile, allowing `LUP` to achieve a consistent 25 FPS. This performance conveniences the standard capture rate of surgical endoscopes, ensuring a fluid, real-time experience for the surgeon without the environmental or infrastructure *tax* of a high-end GPU cluster.

Beyond efficiency, the method integrates uncertainty quantification, which is vital for verifying data reliability in high-stakes environments. By providing a measure of confidence, the system can autonomously filter and validate poses before they are passed to registration or navigation modules. Ultimately, by relying on general geometric assumptions rather than purely data-driven black boxes, LUP is broadly applicable in standard clinical settings without the need for additional instrumentation or energy-intensive hardware.

In the end, Table 3.1 demonstrates the summary of the most relevant works in this field.

Table 3.1: Related work in image-based surgical tool pose estimation.

Reference	Tool	#DoF	~~Init.~~	1 img.	FPS	~~Mrk.~~
[Sarikaya *et al.* 2017]	Surgical*	2**	✓	✓	10	✓
[Laina *et al.* 2017]	Surgical*	2**	✓	✓	18	✓
[Kurmann *et al.* 2017]	Surgical*	2**	✓	✓	9	✓
[Liu *et al.* 2022]	Surgical*	2**	✓	✓	21.6	✓
[Ye *et al.* 2016]	Surgical*	6	×	×	29	✓‡
[Allan *et al.* 2015]	Surgical*	6	×	×	0.9	✓
[Allan *et al.* 2018]	Surgical*	6	×	×	3	✓
[Doignon & Mathelin 2007]	Cylinder	4	✓	✓	×	✓
[Zenteno *et al.* 2021]	Fiberscope	5	×	×	†	✓
LUP (proposed)	LUS	5	✓	✓	25	✓

* Passive tools not comprising sensors

** In 2D image space

† Just mentioned "real-time" and FPS not reported

‡ Uses robot encoders as additional sensor

3.2.2 Sphere and Cylinder Reconstruction

Both cylinders and spheres are quadric surfaces and can, in principle, be reconstructed from image observations. Although multi-view approaches exist [Cross & Zisserman 1998], we focus here on the single-image setting. Existing single-view reconstruction methods typically follow a two-stage pipeline for both shapes: (1) detecting the occluding contour, and (2) reconstructing the 3D shape from that contour. For a cylinder, the occluding contour consists of two parallel lines [Kallasi *et al.* 2014, Gummeson *et al.* 2022, Gummeson & Oskarsson 2023], from which the cylinder axis can be recovered, and its depth estimated if the radius is known. For a sphere, the occluding contour is an ellipse, which can be estimated using [Fitzgibbon *et al.* 1999]; the sphere center is then recovered from the ellipse parameters when the radius is known [Wong *et al.* 2008].

The primary limitation of these methods lies in their computational cost. In particular, reconstructing a sphere from a known radius requires at least 5 contour points, which makes robust estimation frameworks such as RANdom SAmple Consensus (RANSAC) prohibitively expensive, since the number of random samples grows combinatorially with the minimal set size. Two recent methods [Toth & Hajder 2023, Ozgur *et al.* 2026] address this issue by estimating the sphere parameters directly from the silhouette without explicit ellipse fitting. Their minimal case uses only 3 points, thereby significantly accelerating RANSAC.

In contrast, we model the LUS head as a spherocylinder, a hemisphere attached to a cylindrical shaft, and exploit this composite structure. We introduce a novel minimal formulation for hemisphere reconstruction from a single view that requires only 2 points on the hemisphere silhouette, while the remaining constraints are provided by the two silhouette lines of the cylindrical portion. This reduction in minimal sample size yields an additional speed-up when integrated into RANSAC, enabling practical real-time performance.

3.3 Methodology

This section begins with the formal statement of the problem. It then provides an overview of the proposed solution pipeline, followed by a detailed description of each component in the subsequent subsections.

3.3.1 Problem Statement

Given a monocular RGB image $\mathcal{I}_{LAP}$ of a calibrated laparoscope and the spherocylinder LUS head radius r, we estimate the pose $\boldsymbol{\xi}$ of the LUS probe seen in $\mathcal{I}_{LAP}$.

3.3.2 Notation and Modeling

We define $\mathbf{s}_i \in \Re^2$, $\mathbf{c}_j \in \Re^2$, and $\mathbf{o}_k \in \Re^2$ as the occluding contour points of the hemisphere, cylinder, and other possible outlier points, respectively. The point indices count up $i = 1 \ldots n \in \mathbb{N}$, $j = 1 \ldots m \in \mathbb{N}$ and $k = 1 \ldots t \in \mathbb{N}$. The set of unclassified image contour points is then $\mathcal{P} = \{\mathbf{s}_i,\, \mathbf{c}_j,\, \mathbf{o}_k \,|\, \forall\, i, j, k\,\}$. The laparoscope's intrinsic parameter matrix is denoted as $\mathbf{K} \in \Re^{3\times3}$ and the lens distortion parameters with the set $\mathcal{D}$. These parameters are used to project a 3D point in the camera coordinate frame ${}^{c}\mathbf{X} \in \Re^3$ to an image point $\mathbf{x} \in \Re^2$. Equivalently, $\bar{\mathbf{x}} \propto \mathbf{K}\,{}^{c}\mathbf{X}$ where $\bar{\mathbf{x}} = [\mathbf{x}^\top,\, 1]^\top$ is the image point in homogeneous coordinates. The left superscript c indicates the camera coordinate frame. A line $\boldsymbol{\ell} \in \Re^3$ in Hesse normal form on the image plane holds $\boldsymbol{\ell}^\top\, \bar{\mathbf{x}} = 0$ for any of its points $\bar{\mathbf{x}}$. The projection of line $\boldsymbol{\ell} \in \Re^3$ is a plane whose unit normal vector is ${}^{c}\underline{\mathbf{m}} = \mathbf{K}^\top \boldsymbol{\ell} \,/\, \|\, \mathbf{K}^\top \boldsymbol{\ell}\,\|$. The occluding contour of the LUS head's cylinder forms a pair of image lines denoted $\boldsymbol{\ell}_+$ and $\boldsymbol{\ell}_-$, oriented outwards from the cylinder. The normals of the projection planes of these lines are denoted as ${}^{c}\underline{\mathbf{m}}_+$ and ${}^{c}\underline{\mathbf{m}}_-$. The projection of an image point $\mathbf{x} \in \Re^2$ is a ray that starts from the optical center and passes through this image point. We denote the direction of the ray with a unit vector ${}^c\underline{\mathbf{x}} = {}^c\mathbf{x}/\|\,{}^c\mathbf{x}\,\|$ where ${}^c\mathbf{x} = \mathbf{K}^{-1}\,\bar{\mathbf{x}}$. The hemisphere center of the spherocylinder is denoted as ${}^c\mathbf{t}$ and the direction of the ray which passes through it as ${}^c\underline{\mathbf{t}}$. We denote the cylinder axis direction of the spherocylinder as ${}^c\underline{\mathbf{u}}$. Finally, we assemble the LUS pose into a vector $\boldsymbol{\xi} = [\,{}^c\mathbf{t}^\top,\, {}^c\underline{\mathbf{h}}^\top\,]^\top \in \Re^{6\times 1}$ with $\|{}^c\underline{\mathbf{h}}\| = 1$. Fig. 3.1 shows the spherocylindrical LUS head model.

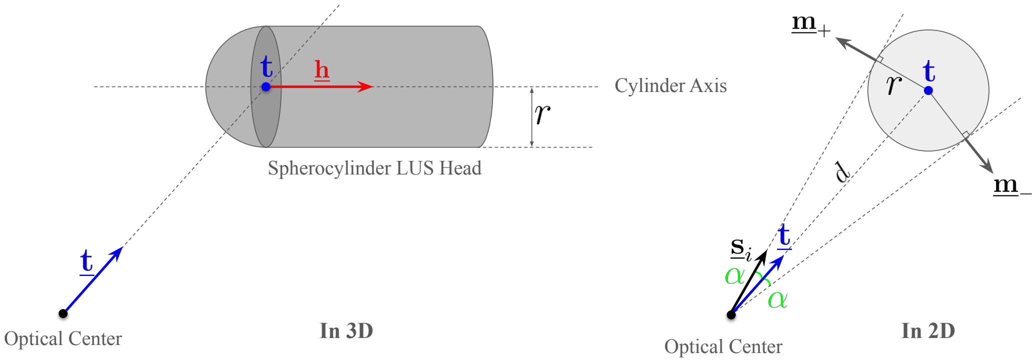


Figure 3.1: Proposed LUS head geometry from two viewpoints.

3.3.3 Solution Pipeline

As illustrated in Fig. 3.2, the proposed pipeline consists of two principal stages: *(i)* segmentation of the LUS probe and extraction of its contour, and *(ii)* pose estimation from unclassified contour points, which are described in detail in the ensuing two sections. Each stage constitutes a challenging problem in its own right.

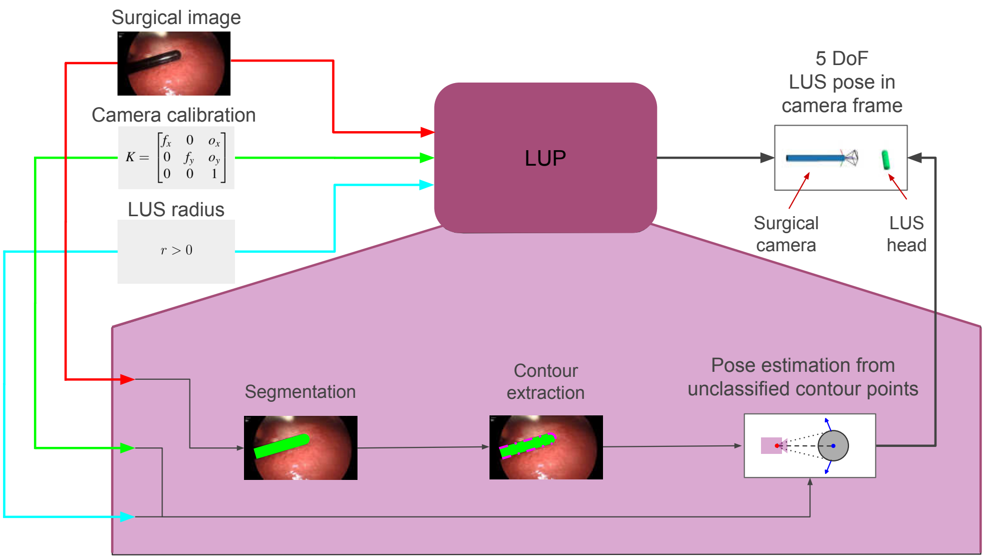


Figure 3.2: LUP pipeline.

3.3.4 LUS Probe Segmentation and Contour Extraction

Dataset Preparation and Data Augmentation From the dataset described in Sec. 2.3.1.2, comprising images acquired from 12 surgical procedures, data were partitioned at the surgery level to ensure complete separation between training, validation, and test samples. Specifically, data from 6 surgeries were assigned to the training set, 1 surgery to the validation set, and the remaining 5 surgeries to the test set. This split was designed to prevent data leakage across subsets and to evaluate the generalization capability of the model on previously unseen procedures.

To improve the robustness of the network and reduce overfitting, several data augmentation strategies were applied online within the data loader during training. Photometric augmentations were employed to encourage the model to learn shape- and structure-based representations of the LUS probe rather than relying solely on color or illumination cues. These augmentations included color jittering, random channel permutation, additive Gaussian noise, posterization, and solarization. In addition, non-photometric geometric transformations were applied, including random horizontal and vertical flips, rotations, and cropping, in order to increase invariance to viewpoint and spatial variation.

Furthermore, a dedicated zoom-out augmentation module was specifically developed for the MIS setting. Since conventional laparoscopic imaging inherently exhibits a circular field of view, this augmentation simulates realistic zoom-out effects by generating and applying a circular mask consistent with the visual characteristics of standard laparoscopic cameras. This operation was introduced to better reflect intraoperative imaging conditions and to improve the robustness of the model to variations in scale and framing.

Model Selection and Training We trained and evaluated several state-of-the-art semantic segmentation architectures, including UNet [Iglovikov *et al.* 2017, Ronneberger *et al.* 2015], TernausNet [Iglovikov & Shvets 2018], and LinkNet [Chaurasia & Culurciello 2017], with particular consideration given to LinkNet-34, as employed in [Shvets *et al.* 2018]. In addition, FarSeeNet [Zhang & Zhang 2020] was investigated and ultimately selected as the most suitable architecture, providing the best trade-off between segmentation accuracy and computational efficiency for the proposed application.

The final model configuration was based on FarSeeNet with a ResNet-18 backbone and a Cascaded Factorized Atrous Spatial Pyramid Pooling (CF-ASPP) module, implemented in PyTorch. Input images were resized to a resolution of 960×540 pixels (downsampled by a factor of 2). The network was trained for 56 epochs using a batch size of 12 and an initial learning rate of 1.3×10^{-4}. Optimization was performed using the Adam optimizer [Kingma & Ba 2017], with Cross-Entropy loss as the training objective. Model performance was primarily assessed using the Intersection over Union (IoU) metric, which reached to 91% on the test set.

Runtime Optimization To enable efficient deployment in latency-constrained intraoperative settings, the trained segmentation network was optimized using NVIDIA TensorRT, a high-performance inference Software Development Kit (SDK) designed for accelerated deployment on NVIDIA GPUs [NVIDIA 2025b]. TensorRT improves inference performance by transforming the trained network into an optimized execution graph tailored to the target hardware architecture. During engine construction, TensorRT applies several graph-level optimizations, including layer fusion, kernel auto-tuning, and memory access minimization, in order to reduce computational overhead and improve throughput. In particular, compatible operations such as convolution and activation layers can be fused into a single optimized kernel, thereby reducing the number of GPU kernel launches and intermediate memory transfers [NVIDIA 2025a]. TensorRT also supports mixed-precision inference, including FP16 and INT8, allowing reduced memory consumption and faster execution while maintaining acceptable predictive performance through precision-aware optimization and calibration [NVIDIA 2025b, NVIDIA 2025a]. Following training in PyTorch, the network was exported and converted into a TensorRT inference engine, enabling hardware-specific runtime optimization on the target GPU. This deployment step significantly reduced inference latency and improved the real-time suitability of the proposed framework for surgical applications [NVIDIA 2025b].

In the proposed implementation, TensorRT optimization yielded a substantial improvement in inference performance, reducing the per-frame segmentation time from 39 ms to 16 ms [Cherfioui 2025]. This corresponds to a latency reduction of approximately 59% and an effective speed-up of 2.4×, thereby significantly improving the real-time performance of the segmentation pipeline. Such acceleration is particularly important in intraoperative applications, where low-latency processing is required to ensure timely visual feedback and seamless integration into the surgical workflow. The observed performance gain confirms the practical benefit of TensorRT-based deployment for enabling real-time semantic segmentation in MIS environments.

Contour Extraction Following segmentation, the contour points of the LUS probe are extracted from the predicted binary mask using the border-following algorithm proposed by [Suzuki & Abe 1985]. This method performs topological contour tracing directly on the binary segmentation output and efficiently retrieves the ordered boundary points corresponding to the external contour of the segmented probe. Owing to its low computational complexity and pixel-wise border traversal strategy, the algorithm is particularly well-suited for real-time applications, providing fast and reliable contour extraction with minimal computational overhead. In the proposed pipeline, this step enables efficient conversion of the dense segmentation mask into a compact contour representation while preserving the essential structural information of the LUS probe boundary. Furthermore, the ordered nature of the extracted contour points enables reliable and straightforward contour downsampling, which further reduces computational cost and accelerates the subsequent RANSAC-based pose estimation while maintaining geometric consistency along the probe boundary.

3.3.5 Single-shot 5-DoF Pose Estimation from Unclassified Contour Points

We employ a sampling-based robust estimation framework based on minimal solutions.

Minimal solution for hemisphere reconstruction Our proposed solution is formulated from two silhouette points $\{\, \mathbf{s}_1,\, \mathbf{s}_2 \,\}$ sampled on the hemispherical cap, and two silhouette lines, $\{\, \boldsymbol{\ell}_+,\, \boldsymbol{\ell}_- \,\}$ on the cylinder. The image projections of the two silhouette points define viewing rays with directions $\{\, {}^c\underline{\mathbf{s}}_1,\, {}^c\underline{\mathbf{s}}_2 \,\}$ in the camera coordinate system, whereas the two silhouette lines back-project to interpretation planes with normals $\{\, {}^c\underline{\mathbf{m}}_+,\, {}^c\underline{\mathbf{m}}_- \,\}$. These two planes are tangent to both the hemispherical cap and the cylindrical body of the spherocylindrical probe model. Given the known probe radius r, the pair of tangent planes imposes strong geometric constraints on the probe pose. In particular, the center of the hemisphere and the axis of the cylinder are constrained to lie on a 3D line $\mathcal{L}$, which is parallel to the spatial intersection of the two tangent planes and offset according to the known radius. This line fully determines the orientation of the cylindrical axis, but still leaves one translational DoF unresolved for the hemispherical center along $\mathcal{L}$. To resolve this remaining ambiguity, silhouette point constraints are introduced. Since each silhouette point on the hemispherical contour back-projects to a ray that is tangent to the hemispherical surface, each corresponding viewing ray imposes an additional tangency constraint on the hemisphere center. A single tangent ray restricts the center to two feasible positions along $\mathcal{L}$, yielding a two-fold ambiguity. Incorporating the second tangent ray removes this ambiguity and uniquely determines the hemispherical center. Consequently, the full 6-DoF pose of the spherocylindrical probe can be recovered from this minimal set of two tangent lines and two tangent points.

Concretely, any image point's ray direction ${}^c\underline{\mathbf{s}}_i$ from the hemisphere silhouette holds the following relation due to the rotational symmetry:

$$
{}^c\underline{\mathbf{s}}_i^\top \, {}^c\underline{\mathbf{t}} \;=\; \cos\left(\alpha\right) \tag{3.1}
$$

where ${}^c\underline{\mathbf{t}}$ is the direction of the ray passing through the hemisphere center, and α is an unknown angle encoding the hemisphere center depth as $d \;=\; r/\sin\left(\alpha\right)$. Similarly, any image line $\boldsymbol{\ell}$ tangent to the hemispherical silhouette back-projects to an interpretation plane with outward normal ${}^c\underline{\mathbf{m}}$. Owing to the rotational symmetry of the hemispherical surface, this plane is necessarily tangent to the hemisphere in 3D, and its normal therefore satisfies the following geometric relation:

$$
{}^c\underline{\mathbf{m}}^\top \, {}^c\underline{\mathbf{t}} \;=\; \cos\left(\frac{\pi}{2} + \alpha\right) \tag{3.2}
$$

Algorithm 1: `2-point hemisphere fit`

Inputs: *ray directions* $\{{}^c\underline{\mathbf{s}}_1,\ {}^c\underline{\mathbf{s}}_2\}$ *on hemisphere silhouette, plane normals* $\{{}^c\underline{\mathbf{m}}_+,\ {}^c\underline{\mathbf{m}}_-\}$ *on cylinder silhouette, spherocylinder radius* r.

Output: *Hemisphere center* ${}^c\mathbf{t} \in \Re^3$.

```
/* Design matrix                                   */
```
1 $\mathbf{Q} = [{}^c\underline{\mathbf{s}}_1 +^c \underline{\mathbf{m}}_+,\ {}^c\underline{\mathbf{s}}_1 +^c \underline{\mathbf{m}}_-,\ {}^c\underline{\mathbf{s}}_2 +^c \underline{\mathbf{m}}_+]_{3\times3}$
```
/* Ray direction to hemisphere center              */
```
2 ${}^c\underline{\mathbf{t}} =^c \mathbf{t}/||{}^c\mathbf{t}||$ where ${}^c\mathbf{t} = \mathbf{Q}^{-\top}\,\mathbf{1}_{3\times1}$
```
/* Depth                                           */
```
3 $d = r/\sin\left(0.5\arcsin\left(1 - 1/||{}^c\mathbf{t}||^2\right)\right)$
```
/* Hemisphere center                               */
```
4 ${}^c\mathbf{t} = d\,{}^c\underline{\mathbf{t}}$

Using equations (3.1) and (3.2), the following minimal linear system can be written:

$$\underbrace{\left[{}^c\underline{\mathbf{s}}_1 +^c \underline{\mathbf{m}}_+,\ {}^c\underline{\mathbf{s}}_1 +^c \underline{\mathbf{m}}_-,\ {}^c\underline{\mathbf{s}}_2 +^c \underline{\mathbf{m}}_+\right]}_{\mathbf{Q}_{3\times3}}^{\top}\,{}^c\underline{\mathbf{t}} \;=\; \lambda\,\mathbf{1}_{3\times1} \tag{3.3}$$

where $\lambda = \cos(\alpha) + \cos\left(\frac{\pi}{2} + \alpha\right)$ and $\mathbf{Q}_{3\times3}$ is the minimal design matrix. Solving (3.3) for ${}^c\underline{\mathbf{t}}$ and α leads to the hemisphere center ${}^c\mathbf{t}$, as shown in Algorithm 1. An n-point overdetermined case with $n \geqslant 3$ has the design matrix:

$$\mathbf{Q}_{3\times2n} = \left[{}^c\underline{\mathbf{s}}_1 +^c \underline{\mathbf{m}}_+,\ {}^c\underline{\mathbf{s}}_1 +^c \underline{\mathbf{m}}_-,\ \cdots,\ {}^c\underline{\mathbf{s}}_n +^c \underline{\mathbf{m}}_-\right] \tag{3.4}$$

and solved in the least-squares sense using the pseudo-inverse $\left(\mathbf{Q}^\top\right)^\dagger$ for the scaled ray direction vector as:

$${}^c\mathbf{t} = \left(\mathbf{Q}^\top\right)^\dagger\,\mathbf{1}_{2n\times1} \tag{3.5}$$

The n-point hemisphere reconstruction is obtained by replacing the expressions in lines 1 and 2 of Algorithm 1 with equations (3.4) and (3.5), respectively.

Principal algorithm The complete LUP pipeline is summarized in Algorithm 2. The procedure begins with *LUS probe segmentation and contour extraction* in line 1, where the probe mask is predicted and converted into an ordered contour representation. In line 2, the extracted contour points $\mathcal{P}$ are corrected for lens distortion, yielding undistorted contour observations suitable for geometric processing.

The main estimation stage is carried out in lines 3–10, where probe pose estimation and contour classification are performed jointly through a sequence of robust RANSAC-based fitting stages. Each RANSAC instance iteratively applies a minimal-solution estimator to randomly sampled subsets of contour observations and terminates with the model that maximizes the consensus set. After convergence, the inlier set associated with the best-fit model is classified and removed from $\mathcal{P}$. As a result, each subsequent RANSAC stage operates on a progressively smaller set of

Algorithm 2: `LUP` – Laparoscopic Ultrasound Pose

Inputs: *Laparoscopic image* $\mathcal{I}_{LAP}$*, spherocylinder radius* r*, camera intrinsics* $\{\mathbf{K}, \mathcal{D}\}$

Output: *LUS pose* ξ_{LUS}

1 $\mathcal{P} = \texttt{ExtractContour}(\,\texttt{SegmentLUS}(\,\mathcal{I}_{LAP}\,)\,)$

2 $\mathcal{P} = \texttt{UndistortContour}(\mathcal{P},\, \mathcal{D}\,)$

3 $\{\boldsymbol{\ell}_1, \mathcal{P}_{\boldsymbol{\ell}_1}\} = \texttt{RANSACLine}(\mathcal{P})$ where $\mathcal{P}_{\boldsymbol{\ell}_1} \subset \mathcal{P}$

4 $\mathcal{P} = \texttt{RemovePoints}(\mathcal{P}, \mathcal{P}_{\boldsymbol{\ell}_1})$

5 $\{\boldsymbol{\ell}_2, \mathcal{P}_{\boldsymbol{\ell}_2}\} = \texttt{RANSACLine}(\mathcal{P})$ where $\mathcal{P}_{\boldsymbol{\ell}_2} \subset \mathcal{P}$

6 $\mathcal{P} = \texttt{RemovePoints}(\mathcal{P}, \mathcal{P}_{\boldsymbol{\ell}_2})$

7 $\{\boldsymbol{\ell}_+, \boldsymbol{\ell}_-\} = \texttt{OutwardFromCylinder}(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2)$

8 $\mathcal{P} = \texttt{RemovePointsOutOfCylinder}(\mathcal{P}, \boldsymbol{\ell}_+, \boldsymbol{\ell}_-)$

9 $\{{}^c\mathbf{t}, \mathcal{P}_h\} = \texttt{RANSACHemisphere}(\mathcal{P}, \boldsymbol{\ell}_+, \boldsymbol{\ell}_-, \mathbf{K}, r)$

10 ${}^c\underline{\mathbf{h}} = \texttt{ComputeCylinderDirection}(\boldsymbol{\ell}_+, \boldsymbol{\ell}_-, \mathbf{K})$

11 $\boldsymbol{\xi} \;=\; [\,{}^c\mathbf{t}^\top,\; {}^c\underline{\mathbf{h}}^\top\,]^\top$

candidate contour points, reducing computational cost and accelerating the overall estimation process.

More specifically, line 3 applies RANSAC to fit an image line to the contour set $\mathcal{P}$ and classifies its consensus set as one of the two projected cylinder edges. In line 4, the corresponding inlier points are removed from $\mathcal{P}$. Line 5 then fits a second image line to the reduced contour set and classifies its consensus set as the second cylinder edge, after which line 6 removes these associated points. Once both cylindrical boundaries have been identified, line 7 orients the fitted image lines such that their normals point outward from the cylinder silhouette, ensuring geometric consistency for subsequent pose recovery.

In line 8, contour points lying outside the inferred cylindrical silhouette are rejected, thereby isolating the subset of points corresponding to the hemispherical cap. Line 9 then applies a third RANSAC stage to this reduced set in order to estimate the hemispherical component. This stage employs the proposed 2-point minimal solution described in Algorithm 1 to generate candidate hypotheses and terminates with a least-squares refinement using all inliers in the final consensus set. Once the hemisphere center has been robustly estimated, line 10 computes the cylinder direction vector ${}^c\underline{\mathbf{h}}$ from the two fitted cylinder silhouette lines, selecting its orientation according to the geometric convention illustrated in Fig. 3.1. Finally, line 11 assembles the complete probe pose from the estimated hemisphere center and cylinder direction, yielding the 5-DoF pose of the LUS probe.

3.4 Experimental Results

We report a quantitative sensitivity analysis of `LUP` using simulated, phantom, and ex vivo data, followed by a quantitative and a qualitative evaluation on clinical datasets acquired during real surgical procedures.

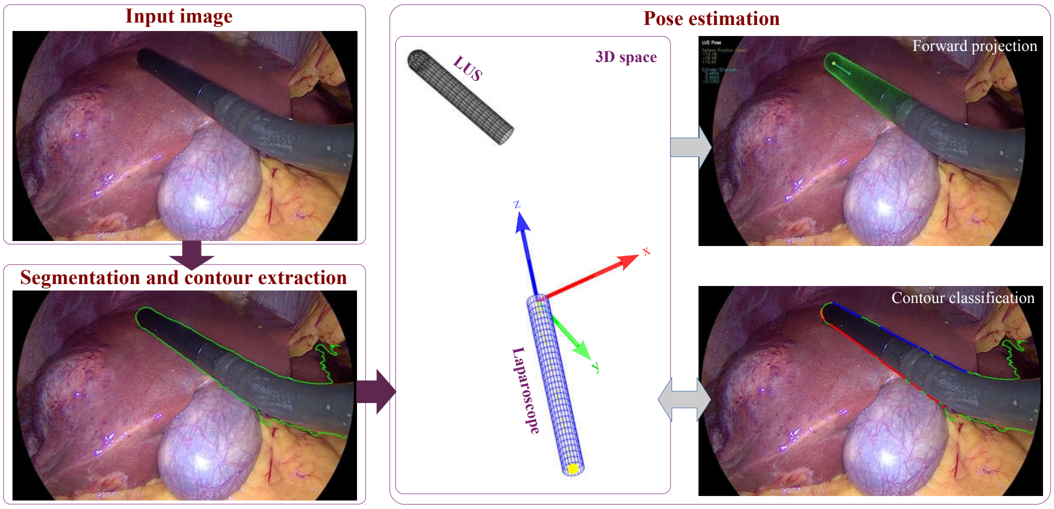


Figure 3.3: `LUP` successfully estimated the LUS probe pose from a single laparoscopic image while classifying contour points into the hemisphere (orange), cylinder edges (blue and red), and outliers (green). The forward-projected probe model achieved a root mean squared error (RMSE) of 1.18 pixels, consistent with the assumed Gaussian measurement noise ($\sigma = 1.25$ pixels).

3.4.1 Sensitivity Analysis

A Monte Carlo simulation framework was used to quantitatively evaluate the sensitivity and robustness of the proposed method under controlled conditions. First, a set of GT LUS probe poses was generated, and the corresponding occluding contours were synthetically projected using the intrinsic calibration parameters of a real laparoscope, with an image resolution of 1920×1080 pixels and a focal length of 1235 pixels. To simulate realistic observation uncertainty, the projected contours were perturbed using multiple pre-generated noise models designed to emulate different sources of contour degradation and localization error. The perturbed contour observations were then used as input to the proposed pose estimation pipeline in order to recover the LUS probe pose. Finally, estimation accuracy was quantified by measuring the discrepancy between the recovered pose and the corresponding GT pose, enabling systematic analysis of the method's sensitivity to contour perturbations.

The inputs to Algorithm 1 consist of two tangent image lines and a set of tip pixels sampled from the hemispherical silhouette. To evaluate the sensitivity of the proposed minimal solver under realistic image uncertainty, these inputs were perturbed using noise distributions derived from real laparoscopic segmentation data, obtained by comparing manual annotations and neural network predictions on clinical images.

For the line noise model, image lines were independently fitted to the manually annotated contours and to the corresponding neural network segmentations. Each line was then represented by the normal vector of its back-projected interpretation plane, expressed in spherical coordinates following the ISO 80000-2:2019 convention. This representation parameterizes the line uncertainty in terms of the angular components θ and ϕ. The resulting error distributions were empirically found to be well approximated by zero-mean Gaussian distributions, with standard deviations of $\sigma_\theta = 1.5 \times 10^{-4}$ rad and $\sigma_\phi = 1.2 \times 10^{-3}$ rad.

A similar procedure was used to construct the tip-pixel noise model. First, hemisphere inlier pixels were extracted from both the manual and neural network segmentations. Using the manually segmented inliers, the reconstructed hemispherical cap was projected into the image plane, yielding a reference ellipse. The Euclidean distances between the neural network hemisphere inliers and this ellipse were then computed to form the tip-pixel error set. Although the resulting distribution exhibited first- and second-order moments of $\mu = 3$ pixels and $\sigma = 2.5$ pixels, respectively, its shape deviated significantly from a Gaussian distribution. Therefore, the empirical samples were retained directly to define a non-parametric noise model. Since this noise is derived from imperfect segmentation output, it represents a conservative estimate of contour uncertainty; lower noise levels are expected under improved segmentation quality.

Estimation error metrics are retrieved from the five pose error metrics defined in Sec. 2.4.1: *in-plane lateral*, *in-plane depth*, and *out-of-plane* translation errors, together with *in-plane* and *out-of-plane* rotational errors.

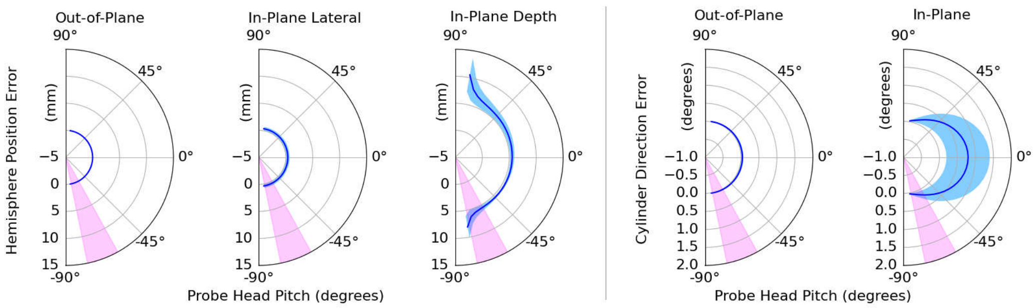


Figure 3.4: Mean absolute errors and standard deviations versus LUS head's pitch in the laparoscope coordinate frame. The rotation axis indicates the angle between the LUS and the xy plane of the laparoscope. The magenta-shaded regions show the LUS head pitch observed in the Surgery 5 video from Fig. 3.9.

The first sensitivity analysis investigated the influence of LUS head orientation. In this experiment, the probe tip was positioned at the image center at a depth of 10, cm from the laparoscope, while the LUS head was rotated about the laparoscope y-axis according to the conventional camera coordinate system shown in Fig. 3.3.

For each orientation, the Monte Carlo simulation was repeated 10,000 times with randomized noise perturbations. In each trial, the projected ellipse corresponding to the hemispherical tip was uniformly sampled at 50 points, from which 10 were randomly selected as input observations. In real data, the average number of hemisphere inliers is approximately 30 points. The results, shown in Fig. 3.4, indicate that the estimated LUS head orientation is highly stable, with negligible rotational error and negligible out-of-plane translation error of the reconstructed hemisphere center. The in-plane lateral translation error remains sub-millimetric across all tested orientations. In contrast, the in-plane depth error is more pronounced and increases with larger LUS head pitch angles. In robotic surgical settings, where both the laparoscope and LUS probe are robotically actuated, these results suggest that additional relative pose constraints could further improve the accuracy of `LUP`.

The second analysis examined sensitivity to probe depth. Since varying other probe orientations did not yield substantially different geometric conditions, depth was isolated as the dominant translational factor influencing estimation accuracy. The simulation setup remained identical to the previous experiment, except that the probe depth was varied over a practical laparoscopic range, and the resulting errors were averaged over all pitch angles treated as a random variable. As shown in Fig. 3.5, both the in-plane depth and in-plane lateral uncertainties of the estimated hemisphere center increase with distance from the camera, consistent with the expected degradation in depth observability under perspective projection.

The third analysis evaluated sensitivity to increasing contour noise. This experiment followed the same configuration as the orientation analysis in Fig. 3.4, except that the sampled noise was scaled by a variable gain factor. The results, summarized in Fig. 3.6, were averaged across all pitch angles. As expected, the estimation error increases approximately linearly with the noise magnitude, indicating that the proposed method exhibits stable and predictable error propagation under increasing observation uncertainty.

Finally, the influence of the number of hemisphere inliers on reconstruction accuracy was analyzed and compared with the recent method proposed in [Toth & Hajder 2023]. As shown in Fig. 3.7, the reconstruction error increases as the number of hemisphere inliers decreases, confirming the expected reduction in geometric constraint. However, beyond approximately 12 inlier points, additional observations provide only marginal accuracy improvement. Compared with [Toth & Hajder 2023], `LUP` achieves superior hemisphere position estimation both in terms of reconstruction accuracy and in the minimum number of silhouette points required for stable estimation.

3.4.2 Phantom Study

Using the phantom dataset described in Sec. 2.3.4, together with the error metrics defined in Sec. 2.4.1 and the GT uncertainty quantification presented in Sec. 2.4.2, a one-minute video sequence was acquired for each of the eight phantoms. Since the visual characteristics of the phantom acquisition environment differ substantially from

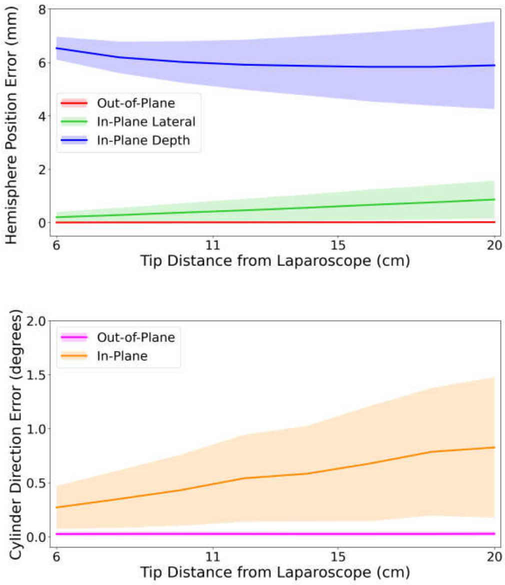


Figure 3.5: LUP's accuracy versus depths.

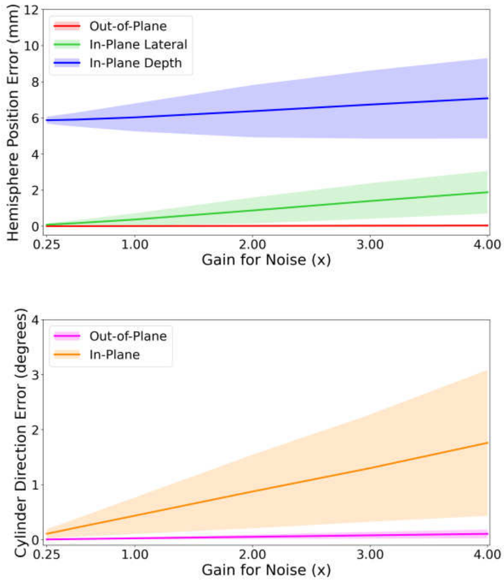


Figure 3.6: LUP's accuracy versus noise levels.

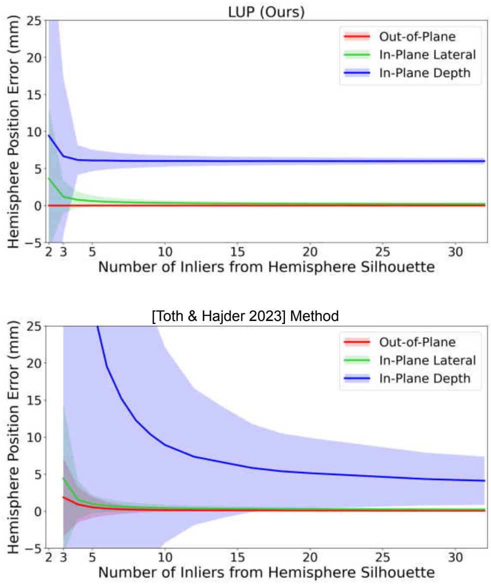


Figure 3.7: LUP and [Toth & Hajder 2023] hemisphere reconstructions versus number of inliers in hemisphere silhouette.

those of the intraoperative surgical scene, the segmentation model trained on surgical data was not directly applied in this experiment. Instead, probe segmentation was obtained using the segmentation-tracking capabilities of SAM3 [Carion *et al.* 2026], initialized with a manual segmentation in the first frame of each sequence.

The remaining stages of the proposed `LUP` pipeline were then applied without modification. In particular, contour extraction, downsampling, and subsequent pose estimation were performed as described previously. The resulting estimated probe poses were compared against the corresponding GT trajectories, and performance was quantified using the evaluation metrics defined earlier. The quantitative results of this phantom-based validation are reported in Table 3.2.

Table 3.2: Average 5-DoF Pose Estimation Error on Phantoms

Translation (mm)			Rotation (Deg)	
Out-of-Plane	In-Plane Lateral	In-Plane Depth	Out-of-Plane	In-Plane
0.2	0.5	7.8	0.1	0.7

3.4.3 Ex Vivo Study

For ex vivo validation, the dataset introduced in Sec. 2.3.2 was evaluated using the performance measures defined in Sec. 2.4.1, while accounting for the reference uncertainty characterized in Sec. 2.4.2. The experimental protocol consisted of four 30-second video acquisitions captured under varying illumination conditions and different laparoscope calibration settings in order to assess the robustness of the proposed method under controlled yet visually challenging imaging conditions.

As illustrated in Fig. 2.4, the visual similarity between the LUS probe and the surrounding organ tissue posed a significant challenge for automated segmentation. In particular, the similarity in color and texture frequently caused segmentation tracking to fail, resulting in unstable or inaccurate probe delineation. Consequently, to ensure a reliable evaluation of the pose estimation stage independent of segmentation failure, all frames in the ex vivo sequences were manually segmented prior to further processing.

The remainder of the proposed `LUP` pipeline was then applied unchanged to these manually segmented masks, and the resulting pose estimates were compared against the corresponding GT trajectories using the predefined evaluation metrics. The quantitative results of the ex vivo validation are summarized in Table 3.3.

Table 3.3: Average 5-DoF Pose Estimation Error on Ex Vivo Organ

Translation (mm)			Rotation (Deg)	
Out-of-Plane	In-Plane Lateral	In-Plane Depth	Out-of-Plane	In-Plane
0.3	0.7	8.1	0.1	0.9

3.4.4 Patient-data Study

Using the available real patient data, both quantitative and qualitative evaluations were conducted to assess the performance of the proposed `LUP` method.

3.4.4.1 Quantitative Study

To further assess the generalizability of the proposed approach, `LUP` was evaluated on the public dataset introduced in Sec. 2.2.1, which provides GT probe poses and was originally published in [Rabbani *et al.* 2022a]. The proposed pipeline was applied to the manually segmented LUS probe heads provided in the dataset, and the resulting pose estimates were compared against the corresponding reference poses reported by the authors. The error for the different available frames for each patient is then averaged, and these quantitative evaluation results are summarized in Table 3.4.

In addition to the numerical analysis, the estimated poses were used to project the LUS probe head model onto the image plane, enabling visual inspection of the geometric alignment between the projected model and the observed probe. One representative frame was selected from each patient sequence for this qualitative assessment, and these examples are presented in Fig. 3.8.

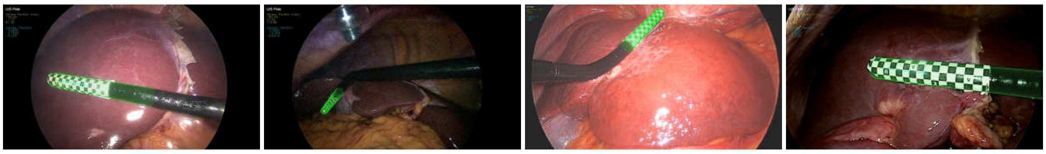


Figure 3.8: `LUP` applied to patients 1 to 4, respectively left to right, from the public dataset with GT of [Rabbani *et al.* 2022a].

	Translation (mm)			Rotation (deg)	
	Out-of-Plane	In-Plane Lateral	In-Plane Depth	Out-of-Plane	In-Plane
P1	0.2	0.8	3.1	0.1	0.8
P2	0.4	1.3	8.2	0.5	2.2
P3	0.4	1.0	5.6	0.4	1.7
P4	0.3	0.7	4.3	0.3	0.7
Avg.	0.3	1.0	5.3	0.3	1.4

Table 3.4: Pose estimation error on the public dataset [Rabbani *et al.* 2022a] averaged over each patient.

3.4.4.2 Qualitative Study

We evaluated `LUP` on laparoscopic image sequences acquired from five distinct liver surgeries, unseen by the segmentation network in the training process. Qualitative results are presented in Fig. 3.3 and 3.9, demonstrating the performance of the proposed method across representative intraoperative scenarios. In all evaluated cases, the projected LUS probe head model, generated from the estimated probe poses, exhibits visually accurate alignment with the corresponding laparoscopic images, indicating precise and stable pose estimation under real surgical conditions.

In this experiment, `LUP` achieved a runtime performance of approximately 25 FPS, enabling near real-time operation. This processing speed is sufficient for the intended clinical application, as the LUS probe is typically manipulated in a slow and controlled manner during intraoperative tumor localization. Consequently, the achieved frame rate provides adequate temporal resolution to ensure smooth visual tracking and responsive pose updates during surgical use.

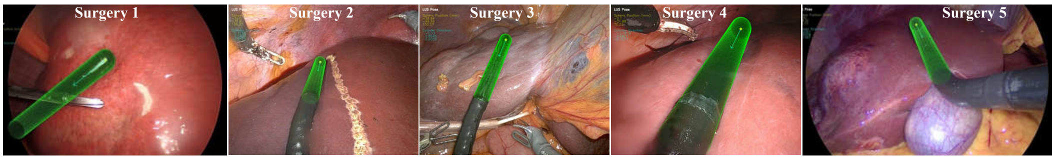


Figure 3.9: Projections of the LUS head model onto the five different surgery images from their estimated poses by `LUP`.

3.5 Conclusion

In this chapter, we introduced `LUP`, the first markerless LUS probe pose estimation framework capable of recovering the 5-DoF pose of the probe directly from standard monocular laparoscopic images. Unlike existing approaches, the proposed method operates without fiducial markers, external tracking hardware, additional sensors, or explicit reinitialization, making it particularly suitable for practical intraoperative deployment. Through extensive experimental evaluation, `LUP` was shown to be robust, accurate, and computationally efficient, achieving pose estimation errors below 1 *cm* while maintaining a runtime of 25 FPS. These characteristics make `LUP` a practical and effective solution for real-time LUS tracking and establish it as a key enabling component for robotic assistance and AR-guided surgical workflows.

Despite these advantages, direct integration of `LUP` into fully constrained AR pipelines remains limited by the unresolved recovery of the sixth-DoF. As demonstrated in this chapter, estimation of the complete 6-DoF probe pose from a single monocular observation of the LUS probe head is fundamentally ill-posed due to the intrinsic geometric symmetry of the probe. Consequently, this missing DoF cannot be recovered from single-frame head visual information alone. Addressing this limitation is essential for complete spatial registration in AR-based surgical guidance. In the following chapter, we therefore extend this formulation by incorporating additional contextual information available in the surgical scene, enabling recovery of the remaining degree of freedom and moving toward full 6-DoF LUS pose estimation.

Chapter 4

Laparoscopic Ultrasound Probe Pose Disambiguation, Filtering, and Imaging Plane Augmentation

Contents

4.1 Introduction

Although MIS has profoundly changed surgical practice by minimizing patient trauma and shortening postoperative recovery, it also introduces significant challenges for the surgical team. One of the most critical issues is the loss of direct palpation and visual assessment of the organ, which markedly limits the surgeon's capacity to accurately localize tumors intraoperatively. To compensate for this limitation, ultrasonography has remained the primary guidance tool for decades [Traynor *et al.* 1988]. However, it is an expensive imaging modality that requires specific radiological expertise and substantial cognitive effort to mentally reconstruct the 3D anatomy from the acquired images. As a consequence of these limitations, more than half of patients undergo non-optimal resections, characterized by insufficient surgical margins, even in tertiary care centers [Viganò *et al.* 2016, Golse *et al.* 2021].

In the presence of missing metastases, namely lesions that have disappeared after chemotherapy while residual tumor tissue persists [Owen *et al.* 2016, Ntourakis *et al.* 2015, Golse *et al.* 2021], no visible scar remains, rendering ultrasonography ineffective for guiding resection or ablation. In such cases, surgeons are often confronted with challenging situations in which resection of the former lesion site is not feasible. Consequently, the effective integration of LUS into the OR relies on accurate probe pose estimation, which constitutes a fundamental prerequisite for advanced assistive solutions such as navigation, guidance, and augmented visualization [Golse *et al.* 2021].

As detailed in Chapter 3, estimating the full pose from a single-shot image of the LUS head is inherently ill-posed when using vision-based methods alone, due to the intrinsic symmetry of the probe along its main axis. This reliance on vision-only constraints is particularly limiting in a sterilized OR environment, where practical considerations restrict the use of additional sensing modalities.

This chapter first explores the use of complementary sources of information to recover the missing DoF in the pose estimation problem. It then incorporates temporal information to enable filtering across successive frames, thereby enhancing the robustness and consistency of the estimated pose. Finally, several approaches to AR visualization are investigated, with the aim of effectively presenting intraoperative information to the surgeon while respecting the constraints of the OR environment.

4.2 Related Work

State-of-the-art approaches for AR-based tumor localization in LLR have yet to converge toward a clinically viable solution that is both practical and non-disruptive to routine surgical workflows. Existing methods can be broadly grouped into sensor-based tracking, fiducial- or marker-based strategies, and learning-based approaches.

Sensor-based techniques commonly employ electromagnetic [Liu *et al.* 2021] or magneto-optic [Feuerstein *et al.* 2007] motion-tracking systems to estimate LUS motion. Although high localization accuracy can be achieved in controlled environments, these systems impose strict line-of-sight requirements, demand meticulous calibration, and necessitate additional operating-room infrastructure. Such requirements are poorly suited to the constrained environment of MIS and can interfere with established surgical workflows. In addition, Electro Magnetic (EM) tracking is susceptible to interference from metallic instruments and field distortions, which can degrade reliability during critical phases of tumor resection.

Fiducial- or marker-based approaches, as in [Pelanis *et al.* 2021, Rabbani *et al.* 2022a], attach physical markers to the LUS to facilitate intraoperative modality registration. Despite their effectiveness in certain scenarios, these markers are prone to migration or detachment, leading to reduced localization accuracy. Their deployment also introduces extra procedural steps, prolongs operation time, and raises sterility-related concerns.

Learning-based methods seek to achieve trackerless AR in LLR through inter-modality registration. In [Montaña-Brown *et al.* 2022], a multi-modal, self-supervised framework is presented to jointly estimate the poses of a laparoscopic camera and an untracked LUS probe using combined video and US data. The approach relies on patient-specific synthetic data generation, employing differentiable rendering to simulate laparoscopic silhouettes and LUS vessel appearances from known poses. A convolutional neural network is trained to regress 6-DoF poses using geometric pose losses together with image re-rendering losses. However, the requirement for extensive patient-specific training increases setup complexity and may limit robustness across heterogeneous patient populations. In contrast, the approach proposed in this work is patient-generic and founded on a principled mathematical model of inter-modality relationships, improving interpretability and allowing step-wise error monitoring and control. [Hadramy *et al.* 2024] introduces a vessel-guided AR framework for liver surgery that unifies heterogeneous intra-operative modalities via a shared vascular centerline representation and integrates biomechanical modeling with DL for real-time non-rigid registration. By leveraging the graph structure of the portal venous tree, the method reports low target registration errors on synthetically generated intraoperative data, even in the presence of partial vessel visibility and segmentation noise. Nevertheless, the approach assumes intraoperative visibility of the root portal vein and sufficient branching, conditions that may not be satisfied under limited acoustic windows, poor contrast, or pathological vascular variations. Furthermore, training and validation are conducted solely on digital-twin–derived data, leaving open questions regarding generalization to real intraoperative settings, where boundary conditions, tissue properties, and force distributions are only coarsely modeled.

By contrast, the markerless and sensor-free AR surgical navigation system presented in this chapter removes the need for external sensors, fiducials, and manual initialization. Direct registration of LUS images to intraoperative MIS video enables real-time, passive, and disruption-free tumor localization. This design preserves surgical workflow integrity while enhancing spatial awareness, allowing seamless visualization of tumor boundaries and critical structures, and thus constitutes a practical and clinically deployable solution for AR-guided minimally invasive surgery.

4.3 Methodology

This section first introduces two distinct methods, each based on different assumptions, for extending the 5-DoF LUS pose estimation provided by the `LUP` method to full 6-DoF pose recovery. Subsequently, the estimated complete probe pose is utilized to perform AR-based guidance by fully registering the LUS modality onto the MIS images as an intraoperative assistive tool.

4.3.1 6th Pose-DoF Estimation via Shaft Constraint and Filtering

Due to the spherocylindrical geometry of the LUS probe head, only five DoF can be estimated from a single-shot visual observation of the head. The remaining sixth-DoF is unobservable as a consequence of the rotational symmetry about the probe head axis. To address this limitation, we introduce the following assumption.

Assumption and Pose Estimation According to Fig. 2.7, we assume that the LUS lever *l2* is kept fixed, such that no lateral rotation of the probe head occurs with respect to the shaft. By construction, this assumption ensures that the LUS imaging plane $\Pi_{\underline{L}}$, shown in blue in Fig. 2.7, remains contained within the plane Π_L defined by the shaft axis line l_s and the head axis line l_h. This assumption is mild and can be readily enforced in practice by instructing the surgeon to refrain from actuating lever *l2* during LUS probe manipulation. We refer to this method as `LUP`*.

Filtering We analyze the LUS probe geometry in order to identify the most appropriate variables for temporal filtering. This choice is motivated by two considerations. First, filtering is restricted to the pose components that are most sensitive to noise. Second, excessive filtering of the full pose is avoided to mitigate the effects of potential inter-variable coupling. As a result, we define the quadruplets $\alpha, \gamma, \theta, \|\mathbf{t}\|$ introduced in Fig. 2.7, which we consider to be sufficiently decoupled. The variable $\|\mathbf{t}\|$ represents an absolute quantity and is estimated independently at each frame. In contrast, the variables α, γ, θ are incremental and are computed as differences between two consecutive frames. These angular increments are expressed in the coordinate system of the subsequent frame. Each variable is filtered independently using a 1-dimensional Unscented Kalman Filter (UKF) with a constant velocity motion model, assuming white noise in both the measurement process and the motion model.

Algorithm Using `LUP`* method for the 6th-DoF estimation and filtering the quadruplets, algorithm 3 is incorporated for our purpose, which we call the method Laparoscopic Augmented Reality from Laparoscopic Ultrasound (LARLUS).

In algorithm 3, line 1 initializes the pose tensor as a zero vector $\mathbf{0}_{6\times1}$ with a history length of one. Line 2 initializes the pose covariance tensor as a 6×6 identity matrix scaled by infinite uncertainty, $\texttt{diag}(\infty)6\times6$, also with a history length of one. Line 3 starts the main processing loop. In line 4, the laparoscopic image $\mathbf{I}_{LAP}$ and the LUS image $\mathbf{I}_{LUS}$ are acquired synchronously. Line 5 estimates the LUS pose $\boldsymbol{\xi}$ and returns the number of observable DoF, given the laparoscope camera intrinsic calibration matrix $\mathbf{K}$, the laparoscopic image $\mathbf{I}_{LAP}$, and the spherocylinder radius r of the LUS probe. This step relies on the `LUP`* algorithm. In line 6, temporal filtering is applied to the quadruplets $\alpha, \gamma, \theta, \|\mathbf{t}\|$ illustrated in Fig. 2.7. These quadruplets are obtained by decoupling the pose history, with relative rotation angles α, γ, and θ computed between consecutive frames. They encode depth variations, which

Algorithm 3: LARLUS

/* Initialize LUS pose history tensor */

1 $\boldsymbol{\psi}_{\blacksquare} \leftarrow \mathbf{0}_{6\times1\times1}$

/* Initialize pose covariance history tensor */

2 $\boldsymbol{\Sigma}_{\blacksquare} \leftarrow \texttt{diag}(\infty)_{6\times6\times1}$

3 **while** *not exit* **do**

  /* Read laparoscope and LUS images */

4   $\{\mathbf{I}_{LAP}, \mathbf{I}_{LUS}\} \leftarrow \texttt{grab\_images}()$

  /* Single-shot pose estimation */

5   $\{\boldsymbol{\psi}, n_{dof}\} \leftarrow \texttt{LUP}^{*}\,(\mathbf{K}, \mathbf{I}_{LAP}, r)$

6   $\{\widehat{\boldsymbol{\psi}}, \widehat{\boldsymbol{\Sigma}}\} \leftarrow \texttt{filtering}\big(\boldsymbol{\psi}_{\blacksquare}, \boldsymbol{\Sigma}_{\blacksquare},\ \boldsymbol{\psi}, n_{dof}\big)$

7   **if** $\texttt{norm}(\widehat{\boldsymbol{\Sigma}}) < \tau$ **then** $\texttt{display}(\texttt{augment}(\mathbf{I}_{LAP}, \mathbf{I}_{LUS}))$ **else** $\texttt{display}(\mathbf{I}_{LAP})$

  /* Stack into the tensors */

8   $\boldsymbol{\psi}_{\blacksquare} \leftarrow \boldsymbol{\psi}_{\blacksquare}.\texttt{add}(\widehat{\boldsymbol{\psi}}),\ \ \boldsymbol{\Sigma}_{\blacksquare} \leftarrow \boldsymbol{\Sigma}_{\blacksquare}.\texttt{add}(\widehat{\boldsymbol{\Sigma}})$

9 **end**

constitute the most sensitive components of the pose estimation. Once filtered, the quadruplets are used to compute the updated LUS pose $\widehat{\boldsymbol{\psi}}$ and its associated covariance $\widehat{\boldsymbol{\Sigma}}$. Line 7 augments the laparoscopic image $\mathbf{I}_{LAP}$ with the LUS image $\mathbf{I}_{LUS}$, provided that the ℓ_2-norm of the pose uncertainty remains below a predefined threshold τ. Finally, line 8 appends the updated LUS pose and covariance to the history tensors for use in the subsequent filtering step.

4.3.2 6th Pose-DoF Estimation via Motion Constraint and Filtering

When either of the assumptions in Sec. 4.3.1 is violated, most commonly due to partial or complete shaft invisibility, the 5-DoF pose estimated from the sphero-cylindrical LUS head can be completed to a full 6-DoF, provided that the following assumptions hold.

Assumption and Pose Estimation The proposed method relies on the following assumptions:

- **Perpendicularity**: The LUS imaging plane is perpendicular to the probe motion trajectory.
- **Motion**: LUS does not only move along the head axis direction or rotate around it.
- **Planarity**: The local surface geometry of the liver is planar.
- **Staticity**: The laparoscope remains static over the last $N + 1$ frames.

Under the last three assumptions over a temporal window of $N+1$ frames, the normal vector of the local liver surface can be estimated as $\underline{\mathbf{g}}$ by fitting a plane to the observed 5-DoF LUS trajectory. Given $\underline{\mathbf{g}}$ and the probe head axis direction $\underline{\mathbf{h}}$, the Perpendicularity assumption enables recovery of the sixth-DoF by computing $\underline{\mathbf{h}} \times \underline{\mathbf{g}}$, which defines the US imaging plane normal at the frame of $(N+1)$.

Filtering After estimating the 6-DoF LUS pose at each frame, temporal filtering is applied to improve robustness and estimation accuracy by enforcing a realistic motion model. To this end, two filtering strategies are implemented:

- **Kalman**: Filtering the quadruplets using a 1-dimensional UKF, as described in Sec. 4.3.1;
- **Graph-based optimization**: Filtering via a Factor Graph representation [Dellaert & Kaess 2017].

The second approach is adopted due to its modularity and flexibility, and it is described in detail in the remainder of this subsection.

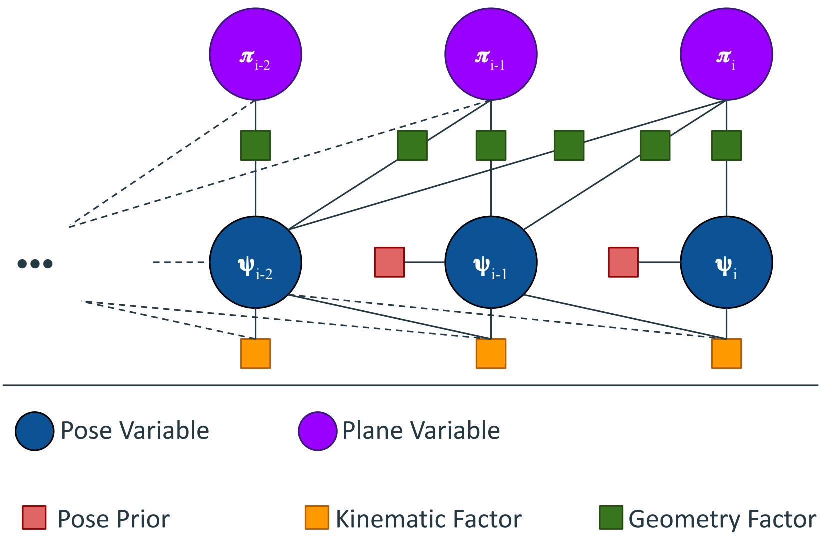


Figure 4.1: Factor Graph representation of the solution.

A schematic overview of the proposed factor graph formulation is illustrated in Fig. 4.1. In this graphical representation, circular nodes denote the optimization variables to be estimated, whereas rectangular nodes represent priors and measurement factors that impose constraints on these variables. Let i denote the current frame index. For each frame, the full 6-DoF pose variable $\boldsymbol{\psi}_i$ is introduced as the primary optimization state. This variable is constrained by a 5-DoF pose prior derived from the `LUP` observation, represented by the red rectangular nodes in Fig. 4.1. Specifically, this prior incorporates the estimated pose parameters $\hat{\boldsymbol{\xi}}$, obtained from

LUP, and constrains the corresponding components of $\boldsymbol{\psi}_i$ according to the formulation in Equation 4.1.

$$f_i^{\text{pose}}(\boldsymbol{\psi}_i, \hat{\boldsymbol{\xi}}_i) = \exp\left(-\frac{1}{2}\left\|\boldsymbol{\psi_i} - \hat{\boldsymbol{\xi}}_i\right\|^2_{\Sigma^{-1}_{\hat{\boldsymbol{\xi}}_i}}\right) \tag{4.1}$$

In Equation 4.1, $\boldsymbol{\psi} = \ln(T)^{\vee}$ (and $T = \exp\left(\boldsymbol{\psi}^{\wedge}\right)$) where $T \in SE(3)$ and $\boldsymbol{\psi} \in \mathbb{R}^6$ (and thus, $\boldsymbol{\psi}^{\wedge} \in \mathfrak{se}(3)$) where $\mathfrak{se}(3)$ is the Lie Algebra associated with $SE(3)$ [Barfoot 2024] and the $(.)^{\wedge}$ operator [Murray 1994, Barfoot 2024] takes elements of $\mathbb{R}^6$ and turns them into elements of $\mathbb{R}^{4\times4}$. We use the inverse of this operation as $(.)^{\vee}$ [Barfoot 2024]. $\hat{\boldsymbol{\xi}} \in \mathbb{R}^6$ is the estimated 5-DoF LUS pose by LUP method, in which the last element, *i.e.*, the ambiguous self rotation, is set to 0. By denoting $\boldsymbol{\delta} \triangleq \boldsymbol{\psi} - \hat{\boldsymbol{\xi}}$ the effect of LUP uncertainty in the pose estimation, $\Sigma_{\hat{\boldsymbol{\xi}}}$, can be explained via Equation 4.2 [Dellaert & Kaess 2017].

$$\|\boldsymbol{\delta}\|^2_{\Sigma} \triangleq \boldsymbol{\delta}^{\top}\Sigma^{-1}\boldsymbol{\delta} = \left(\Sigma^{-1/2}\boldsymbol{\delta}\right)^{\top}\left(\Sigma^{-1/2}\boldsymbol{\delta}\right) = \left\|\Sigma^{-1/2}\boldsymbol{\delta}\right\|^2_2. \tag{4.2}$$

It should be noted that the associated uncertainty of the pose estimation, as the covariance matrix $\Sigma_{\hat{\boldsymbol{\xi}}}$ is represented in the temporal LUS frame, and the uncertainty associated with the self-rotation is theoretically infinite.

In addition to the pose priors, kinematic factors, shown by yellow squares, are incorporated to enforce temporal motion consistency throughout the optimization process. These factors regularize the estimated pose trajectory by introducing dynamic constraints between consecutive states, thereby reducing frame-to-frame jitter and improving temporal stability. Under a simple constant-position motion model, each kinematic factor links two consecutive pose variables, as formulated in Equation 4.3, resulting in a binary temporal constraint. In this formulation, Σ_k is the covariance matrix for this motion model, similar to what is used in UKF.

$$f_i^{\text{kin}}(\boldsymbol{\psi}_i, \boldsymbol{\psi}_{i-1}) = \exp\left(-\frac{1}{2}\left\|\boldsymbol{\psi_i} - \boldsymbol{\psi}_{i-1}\right\|^2_{\Sigma_k^{-1}}\right) \tag{4.3}$$

For more expressive motion modeling, this formulation can be extended to a constant-velocity model, in which the kinematic constraint is defined over three consecutive poses, as described by Equation 4.4. In this case, each kinematic factor forms a ternary constraint that enforces consistency in both position and velocity across adjacent time steps. In this formulation, $\left(2\boldsymbol{\psi}_{i-1} - \boldsymbol{\psi}_{i-2}\right)$ is the pose prediction under the constant-velocity motion model.

$$f_i^{\text{kin}}(\boldsymbol{\psi}_i, \boldsymbol{\psi}_{i-1}, \boldsymbol{\psi}_{i-2}) = \exp\left(-\frac{1}{2}\left\|\boldsymbol{\psi_i} - \left(2\boldsymbol{\psi}_{i-1} - \boldsymbol{\psi}_{i-2}\right)\right\|^2_{\Sigma_k^{-1}}\right) \tag{4.4}$$

In the proposed framework, the constant-velocity model is adopted, as it provides a more realistic approximation of the smooth and continuous motion of the LUS probe during intraoperative manipulation while remaining computationally efficient for real-time optimization.

Plane variables are parameterized as $\boldsymbol{\pi} = [\mathbf{n}, d]$, where $\mathbf{n}$ denotes the plane normal and d is the signed distance parameter, such that the plane is defined by the implicit equation $\mathbf{n}^T\mathbf{x} + d = 0$. Within the proposed factor graph, these plane variables encode the geometric constraints associated with the local tissue surface. The observation function corresponding to the geometry factor, denoted by green rectangles, is formulated in Equation 4.5 and jointly enforces two physically meaningful constraints. First, it penalizes the orthogonal distance between the estimated LUS probe position and the plane, ensuring spatial consistency between the probe and the observed tissue surface. Second, it imposes a perpendicularity constraint between the principal axis of the LUS probe head and the plane normal, reflecting the expected acquisition geometry in which the probe is approximately oriented normal to the tissue surface during contact. Together, these constraints provide a geometrically grounded regularization term that improves pose observability and stabilizes the estimation of the otherwise unconstrained degree of freedom.

$$f_i^{\text{geo}}(\boldsymbol{\psi}_i, \boldsymbol{\pi}_i) = \exp\left(-\frac{1}{2}\left\|\begin{bmatrix}\underline{\mathbf{n}_i}^T\mathbf{x}_i + d \\ \underline{\mathbf{n}_i}^T\underline{\mathbf{h}_i}\end{bmatrix} \in \mathbb{R}^2\right\|^2_{\Sigma_{g_i}^{-1}}\right) \tag{4.5}$$

In Equation 4.5, $\mathbf{x}_i$ is the position element of $\boldsymbol{\psi}_i$ in the Cartesian space, and $\underline{\mathbf{h}_i}$ is the LUS head axis unit direction vector in the Cartesian space. The local plane normal $\underline{\mathbf{n}_i}$ and d are estimated via a plane fitting over the window of the last N 5-DoF observed poses, and Σ_{g_i} is its associated covariance retrieved from the plane fitting error.

Finally, the overall optimization for the pose variables is formulated in Equation 4.6, where the estimation is obtained by jointly minimizing the residuals induced by the pose priors, kinematic factors, and geometric constraints. This cost yields the optimal pose trajectory by balancing direct visual observations with temporal smoothness and scene-consistent geometric regularization.

$$\boldsymbol{\Psi} = \underset{\boldsymbol{\Psi}}{\operatorname{argmin}} \sum_{i=a}^{b} \left(f_i^{\text{pose}} + f_i^{\text{kin}} + f_i^{\text{geo}}\right) \tag{4.6}$$

In Equation 4.6, $\boldsymbol{\Psi} = \{\boldsymbol{\psi}_a, \ldots, \boldsymbol{\psi}_b\}$ where a and b are the arbitrary indices of a starting and ending required window, which in practice b is the index of the last frame and $a = b - W$ in which the window length W is a smaller or equal number to N.

Similarly, the utilized optimization for the plane variables is defined in Equation 4.7, where the plane parameters are estimated by minimizing the residuals associated with the corresponding geometric observations and their initialized value by plane fitting. In this equation, $\boldsymbol{\Pi} = \{\boldsymbol{\pi}_a, \ldots, \boldsymbol{\pi}_b\}$.

$$\boldsymbol{\Pi} = \underset{\boldsymbol{\Pi}}{\operatorname{argmin}} \sum_{i=a}^{b} \left(f_i^{\text{init}} + f_i^{\text{geo}}\right) \tag{4.7}$$

Together, these two coupled optimization problems enable consistent joint refinement of the LUS probe trajectory and the underlying local surface geometry.

Algorithm The complete pose estimation and temporal filtering framework is summarized algorithmically in Algorithm 4, which unifies the proposed observation, geometric reasoning, and optimization steps into a single sequential estimation pipeline. This method is called Motion-based Sixth Pose DoF Recovery (MSPDR) henceforth.

In Algorithm 4, the procedure begins by initializing the frame index to zero. At each iteration of the main loop, synchronized MIS and LUS image streams are acquired and processed jointly. For every incoming frame, the 5-DoF pose of the LUS probe is first estimated using `LUP`, together with its associated uncertainty. These per-frame pose observations are then accumulated over time by stacking the estimated states until a temporal window of length N is reached, thereby providing sufficient historical context for geometric reasoning.

Once pose estimates from the preceding N frames are available, the previously unobservable sixth-DoF can be recovered through RANSAC-based plane fitting applied to the accumulated trajectory. This step exploits the local geometric consistency of the probe motion to infer the missing rotational component and produce a fully constrained 6-DoF pose estimate. The recovered poses are subsequently passed to a temporal filtering stage, where they are refined and regularized to improve robustness and suppress frame-to-frame noise. In parallel, the associated uncertainty is evaluated and propagated to support downstream confidence-aware decision making, particularly for selective image augmentation and visualization.

It should be noted that the filtering function in Line 11 is modular and may be implemented using either a UKF, as described in Sec. 4.3.1, or a factor graph-based optimization scheme, as introduced in Sec. 4.3.2. Furthermore, the window length parameter W determines the number of preceding poses included in the filtering or optimization stage. In practice, W is typically chosen to be smaller than N in order to preserve real-time performance while still leveraging sufficient temporal information for stable estimation.

4.3.3 Augmented Reality Rendering

After recovering the full 6-DoF pose of the LUS probe, three visualization strategies are introduced for intraoperative guidance. The first two strategies visualize only the estimated pose and spatial extent of the LUS imaging plane within the laparoscopic scene, while the third provides direct multimodal augmentation by overlaying the LUS image onto the MIS image. All three visualization modes rely on homography estimation to establish the projective mapping between the LUS image plane and the laparoscopic image. The latter two strategies additionally require image blending in order to generate visually coherent augmentations. A comparative evaluation of these visualization modes is presented in Sec. 4.4.4.

Algorithm 4: `MSPDR` – Motion-based Sixth Pose-DoF Recovery

$i \leftarrow 0$ `// Initialize frame number`
while *not exit* **do**
    `/* Read laparoscope and LUS images */`
    $\{\mathbf{I}_{LAP}, \mathbf{I}_{LUS}\} \leftarrow$ `grab_images()`
    `/* Single-shot 5-DoF pose and uncertainty estimation */`
    $\{\boldsymbol{\xi}, \widehat{\boldsymbol{\Sigma}}\} \leftarrow$ `LUP` $(\mathbf{K}, \mathbf{I}_{LAP}, r)$
    `/* Comparing with window length */`
    if $i < N$ **then**
        $\boldsymbol{\Xi}_i \leftarrow [\boldsymbol{\xi}, 0]^T$
        $\boldsymbol{\Sigma}_i \leftarrow \begin{bmatrix} \widehat{\boldsymbol{\Sigma}} & \mathbf{0}_{5\times1} \\ \mathbf{0}_{1\times5} & \infty \end{bmatrix}$
    end
    else
        `/* Full pose estimation via 6th-DoF recovery */`
        $\{\hat{\psi}, \widehat{\boldsymbol{\Sigma}}\} \leftarrow$ `est_sixth`$\left(\boldsymbol{\xi}, \widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Xi}, \boldsymbol{\Sigma}, N, i\right)$
        $\{\boldsymbol{\Xi}_i, \boldsymbol{\Sigma}_i\} \leftarrow$ `filtering`$\left(\hat{\psi}, \widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Xi}, \boldsymbol{\Sigma}, W, i\right)$
        if *norm* $(\boldsymbol{\Sigma}_i) < \tau$ **then**
            `display(augment(`$\mathbf{I}_{LAP}, \mathbf{I}_{LUS}, \boldsymbol{\Xi}_i$`))`
        end
        else
            `display(`$\mathbf{I}_{LAP}$`)`
        end
    end
    $i \leftarrow i + 1$
end

Homography Estimation Homography estimation requires at least four pairs of corresponding points between the LUS image $\mathbf{I}_{LUS}$ and the laparoscopic image $\mathbf{I}_{LAP}$. In the proposed framework, these correspondences are defined using the four corner points of the LUS imaging plane. The corner points are first expressed in the laparoscope coordinate frame through the estimated 6-DoF probe pose. The top-left corner of the LUS imaging plane is defined at a fixed and known position relative to the LUS probe tip, corresponding to the center of the hemispherical probe head. Once this reference point is localized with respect to the probe geometry, using calibration parameters, the LUS probe CAD model, or direct physical measurements, the remaining three corners are determined directly from the estimated probe pose and the known dimensions of the LUS image plane, namely its width and height as specified by the ultrasound system configuration. These four corner points are then projected into the laparoscopic image plane to establish four-point correspondences, from which the planar homography is computed.

Blending To generate an interpretable multimodal overlay, the LUS and MIS images are blended through a five-stage image compositing process. First, the grayscale LUS image is converted into a green-scale representation in order to enhance perceptual contrast against the native laparoscopic image while preserving anatomical visibility. Second, a spatially varying transparency mask is applied, such that the opacity decreases linearly toward the image boundaries, thereby producing a smoother visual transition and reducing abrupt overlay edges. Third, the resulting green-scale LUS image is warped into the laparoscopic image domain using the estimated homography. Fourth, the warped LUS image is blended with the laparoscopic image using linear alpha compositing to generate the augmented visualization. Finally, the segmented LUS probe, extracted from the laparoscopic image using the segmentation mask predicted by the network, is rendered on top of the blended image. This final occlusion-aware compositing step ensures that the probe head remains visually consistent with the scene geometry and correctly occludes the augmented LUS content, thereby improving depth perception and visual realism in the resulting overlay.

4.4 Experimental Results

Although rigorous validation remains inherently challenging in medical imaging studies, the proposed methods were evaluated through a comprehensive experimental protocol spanning phantom, ex vivo, and clinical settings. This multi-stage validation strategy was designed to assess the proposed framework under progressively more realistic conditions, ranging from controlled laboratory environments to clinically relevant intraoperative scenarios. The UKF-based implementation was realized using the FilterPy library, while the graph-based optimization framework was implemented using iSAM2 [Kaess *et al.* 2011], enabling efficient incremental nonlinear optimization for real-time pose refinement.

The temporal window parameter N used in the motion-based pose estimation stage was empirically set to 200, corresponding to approximately 7 seconds of probe motion. This value was found to provide sufficient temporal context for robust recovery of the missing degree of freedom while maintaining stable performance across all experimental settings. The optimization window parameter W, which defines the number of most recent poses jointly refined at each iteration, was set to 3. This choice provided an effective compromise between temporal consistency and computational efficiency, enabling local trajectory refinement while preserving real-time performance.

In the experiments, LARLUS achieved a runtime performance of approximately 22 FPS, whereas MSPDR operated 15 to 17 FPS, depending on the dataset characteristics. The lower rate of MSPDR is primarily attributed to the additional computational cost introduced by the RANSAC-based plane fitting stage, which increased the overall processing complexity. Nevertheless, both methods maintained real-time performance suitable for intraoperative applications.

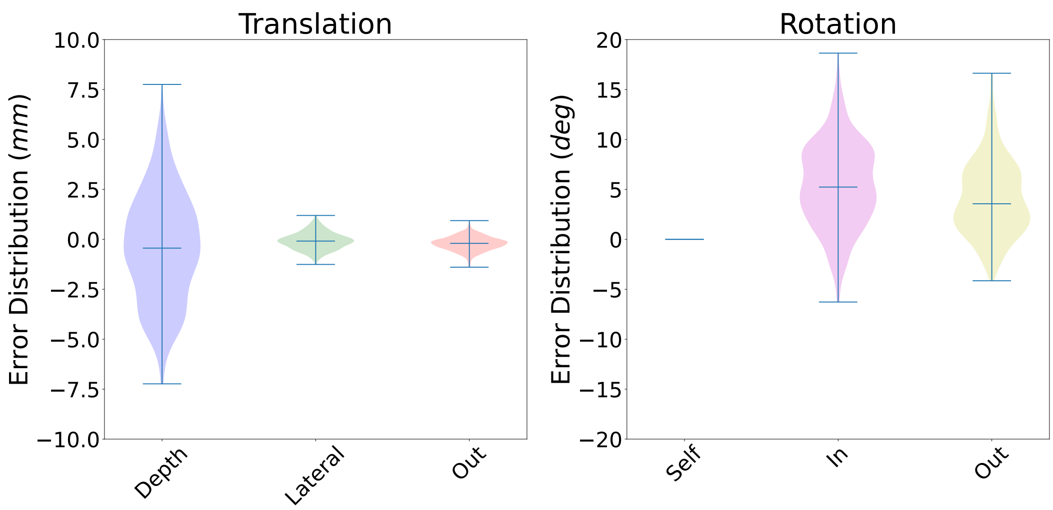


Figure 4.2: LUP errors on the phantom. Note that LUP is a 5-DoF pose estimation method, and the self-rotation value is not estimated.

4.4.1 Phantom Study

For this study, the silicone-based deformable phantom introduced in Sec. 2.3.4, derived from a real patient-specific liver model, was used to evaluate the proposed method under controlled yet anatomically realistic conditions. Owing to its deformable silicone structure, the phantom provides a suitable experimental surrogate for assessing intraoperative tissue interaction and probe contact behavior. When the phantom surface is soaked with water, effective contact between the ultrasound probe and the phantom can be directly verified from the acquired ultrasound images, thereby providing a practical visual indicator of proper acoustic coupling during data acquisition. An ArUco marker attached to the probe was used exclusively to provide GT pose information for quantitative evaluation, and system calibration was performed according to the protocol described in Sec. 2.3.2. The evaluated dataset consists of 500 frames, and image segmentation was performed using SAM3 [Carion *et al.* 2026].

The quantitative performance of the different evaluated methods is summarized in Fig. 4.2, Fig. 4.3, Fig. 4.4, and Fig. 4.5, where the results are reported using the error metrics defined in Sec. 2.4.1. These results provide a comparative assessment of the proposed framework in terms of geometric accuracy and temporal stability under deformable phantom conditions.

Fig. 4.2 presents the error distributions of the baseline LUP method on the phantom dataset, together with their median values shown as violin plots. As a single-shot method relying exclusively on visual observations of the LUS probe head, LUP estimates only 5-DoF and, by construction, cannot recover the probe self-rotation around its principal axis. Consequently, the sixth-DoF remains unobservable in this formulation.

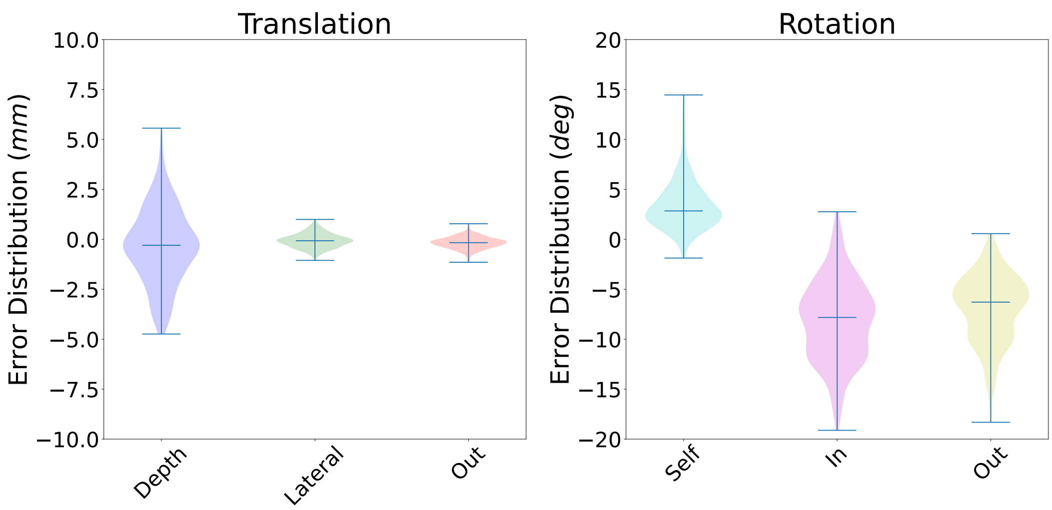


Figure 4.3: LARLUS errors on the phantom.

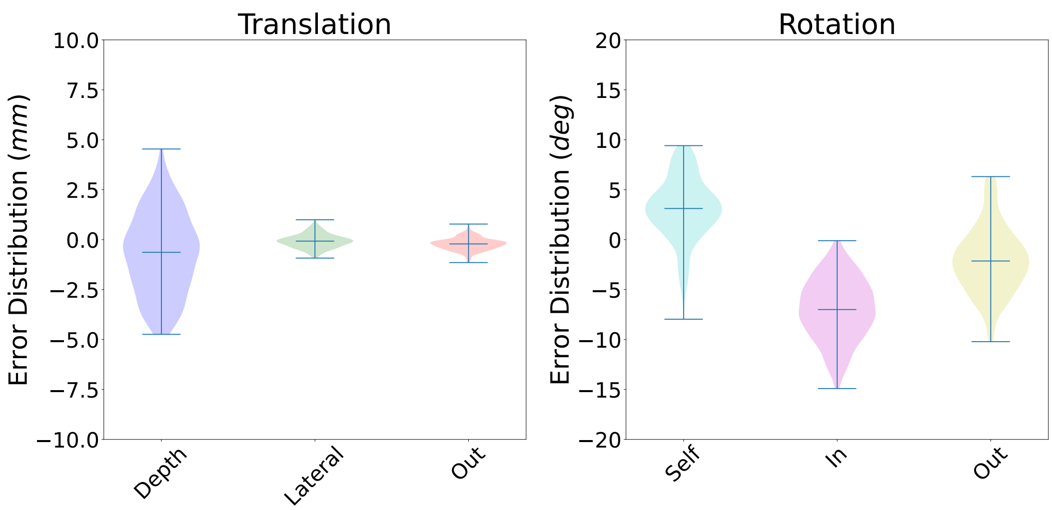


Figure 4.4: MSPDR errors on the phantom with Kalman filter.

By introducing the shaft-based geometric constraint, the LARLUS method resolves this ambiguity and recovers the missing sixth-DoF. In addition, by incorporating inter-frame motion consistency, LARLUS performs temporal filtering of the pose estimates, resulting in improved robustness and reduced estimation error. This improvement is reflected in the narrower error distributions and lower median values observed in Fig. 4.3.

Similarly, the proposed MSPDR framework enables full 6-DoF pose estimation by recovering the previously unobservable rotational component through motion-based geometric reasoning. When combined with Kalman filtering, the resulting error distributions are shown in Fig. 4.4, while the incorporation of graph-based optimization yields the results presented in Fig. 4.5. A comparison of these two

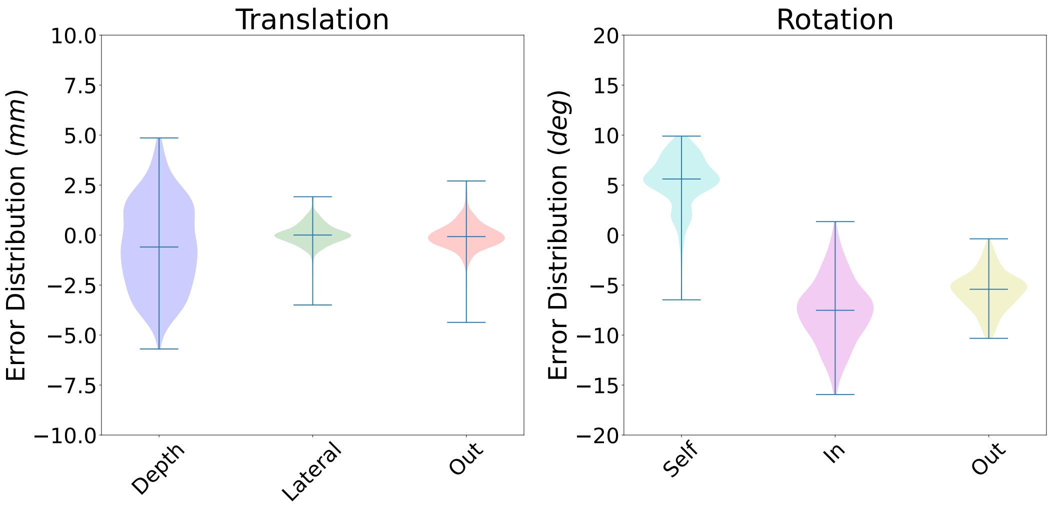


Figure 4.5: MSPDR errors on the phantom with graph-based optimization.

filtering strategies within the MSPDR framework indicates complementary behavior: the Kalman-based formulation provides slightly better positional accuracy, whereas the graph-based optimization achieves superior orientation refinement.

A direct comparison between LARLUS and MSPDR should be interpreted with care, as the two methods rely on different assumptions and are designed for different levels of applicability. Nevertheless, in scenarios where both methods are applicable, as in the present phantom experiment, MSPDR demonstrates lower error variance and improved estimation stability, indicating a more consistent pose recovery under the evaluated conditions.

4.4.2 Ex Vivo Study

For this study, the ex vivo dataset introduced in Sec. 2.3.2 was used for quantitative evaluation. An ArUco marker attached to the LUS probe served exclusively as a source of ground-truth pose information and was used only for performance assessment, not for pose estimation. The evaluated sequence consists of 500 frames. Owing to the low visual contrast between the organ surface and the LUS probe in the laparoscopic images, automatic segmentation proved less reliable in this setting. Therefore, image segmentation was performed manually in order to ensure accurate probe delineation and to reduce segmentation-induced uncertainty. This manual refinement provided more precise contour extraction and improved overall evaluation accuracy compared with the automatic segmentation used in the preceding phantom study in Sec. 4.4.1.

The results of this study are presented in Fig. 4.6, Fig. 4.7, Fig. 4.8, and Fig. 4.9. The comparative trends observed in the phantom experiments presented in Sec. 4.4.1 are likewise preserved in the present ex vivo evaluation. In particular, the relative behavior of the different methods, as well as the performance trade-offs between

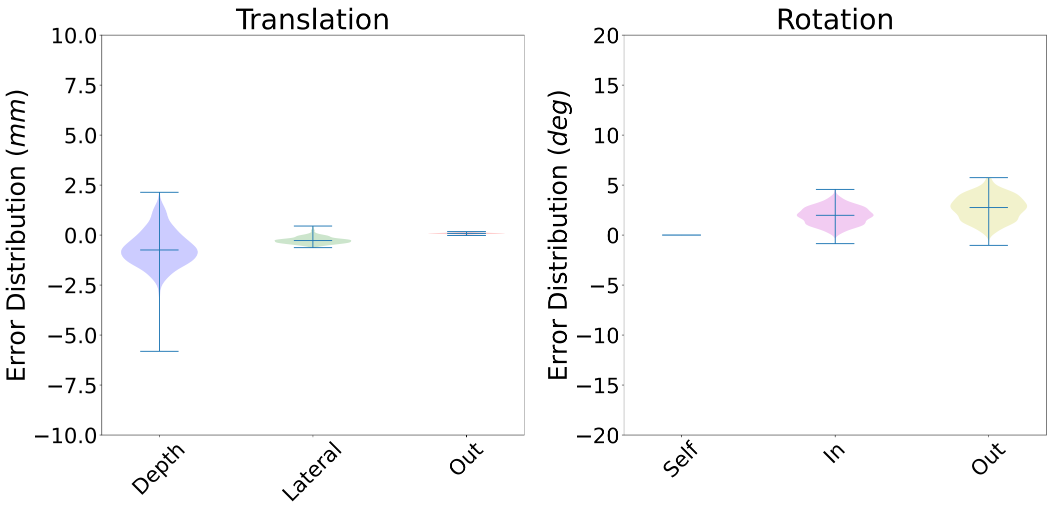


Figure 4.6: LUP errors on the ex vivo organ. Note that LUP is a 5-DoF pose estimation method, and the self-rotation value is not estimated.

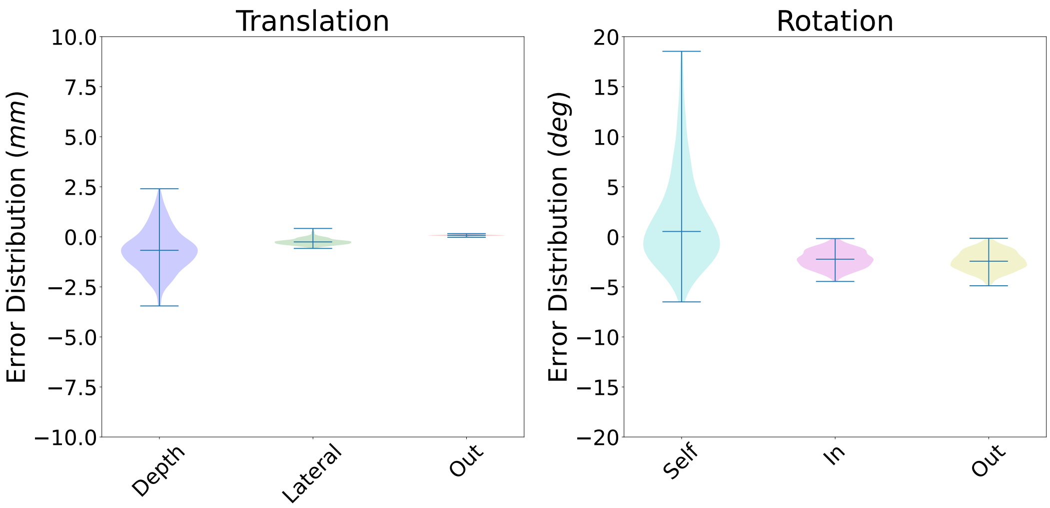


Figure 4.7: LARLUS errors on the ex vivo organ.

filtering strategies, remains consistent with the observations previously discussed. As these results follow the same overall pattern and confirm the conclusions drawn from the phantom study, the detailed comparative analysis is not repeated here.

4.4.3 Patient-data Study

In this study, a dataset consisting of 144-frame image sequences was recorded during a clinical hepatectomy procedure. To maintain temporal continuity and extend the dataset length, the sequence was played forward and backward four times, resulting in a total of $144 \times 4 = 576$ frames. The recorded dataset satisfies the assumptions of

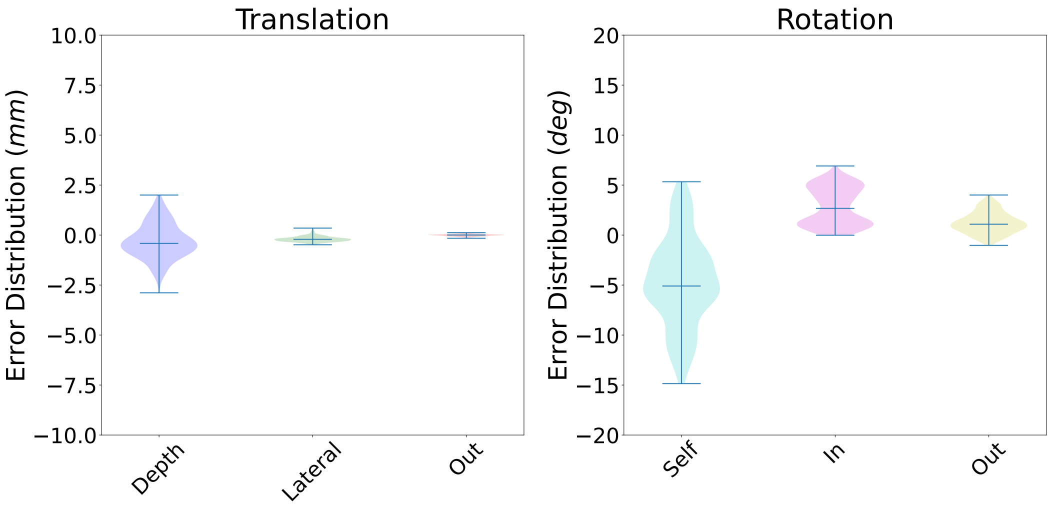


Figure 4.8: MSPDR errors on the ex vivo organ with Kalman filter.

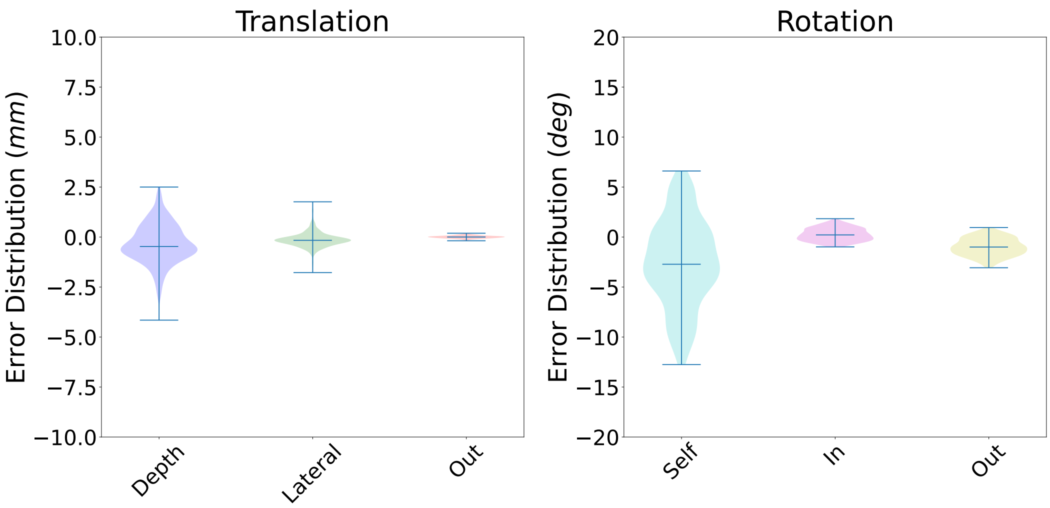


Figure 4.9: MSPDR errors on the ex vivo organ with graph-based optimization.

LARLUS; however, it does not meet the assumptions of the MSPDR method and does not include GT data due to sterility constraints.

For validation, we first evaluate the effect of filtering by examining parameter variations, as shown in Fig. 4.10. As can be seen, the filtered parameters exhibit substantially less fluctuation compared to the unfiltered measurements.

In the next validation step, the effect of parameter updating is examined qualitatively using the video sequence, as illustrated in Fig. 4.11. The upper row shows the LUS imaging planes without filtering, while the lower row shows the same planes after filtering. In each frame, the green contour represents the current configuration at time k, and the white contour represents the configuration at the previous time step $k-1$. It can be observed that, without filtering, even minor movements of the

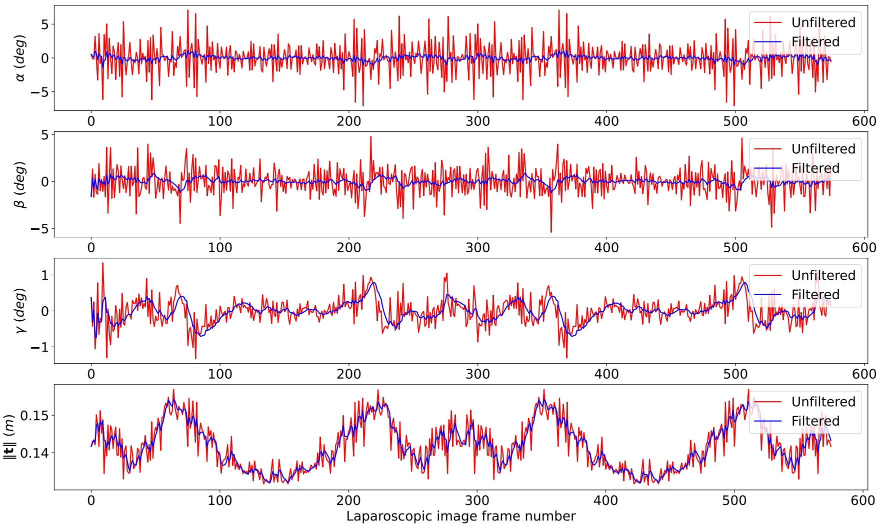


Figure 4.10: Unfiltered and filtered depth-related variables.

LUS probe can cause substantial flickering of the imaging plane due to sensitivity to noise.

Finally, the augmented output is validated by the surgeon, confirming that it introduces no noticeable lag or eye strain from flickering. Additionally, the accuracy of the LUS plane pose is assessed qualitatively based on the surgeon's expertise while reviewing the video. Some of the video frames are brought in Fig. 4.12.

4.4.4 Visualizations

To leverage the estimated pose for visualization, three approaches were implemented and evaluated. The first approach visualizes only the LUS plane with the scales on the MIS frames (Fig. 4.13), offering minimal clutter in the surgical scene. In the second approach (Fig. 4.14), the LUS image is directly augmented onto the surgical scene at the location of the estimated plane. While this reduces the need for the surgeon to switch between MIS and LUS monitors, it increases scene clutter and suffers from reduced visibility when the probe head axis aligns closely with the laparoscope's $\underline{\boldsymbol{z}}$ axis. The third approach (Fig. 4.15) addresses this issue by displaying the original LUS image in a corner of the MIS frame. Although this preserves visibility regardless of probe orientation, it introduces the highest level of scene clutter.

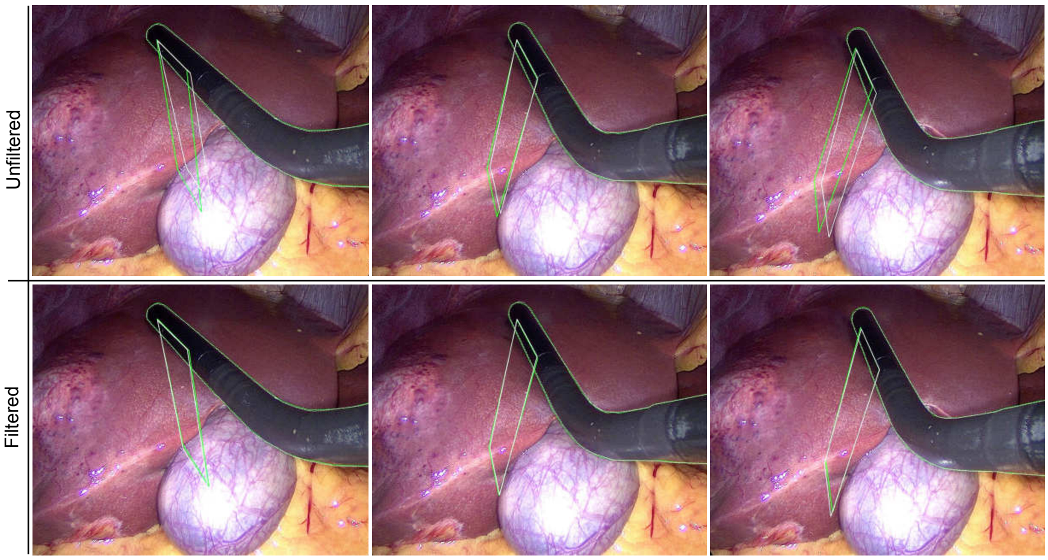


Figure 4.11: The upper row illustrates the estimated LUS imaging planes without temporal filtering, whereas the lower row shows the corresponding results after filtering. In each frame, the green contour represents the current LUS imaging plane configuration at time step k, while the white contour indicates the configuration at the previous time step $k-1$. It can be observed that, without filtering, even minor LUS probe motion leads to noticeable fluctuations in the estimated imaging plane, resulting in visible flickering due to sensitivity to measurement noise. In contrast, temporal filtering yields a more stable and visually consistent plane estimation across consecutive frames.

Figure 4.12: Multiple augmentation instances from `LARLUS`.

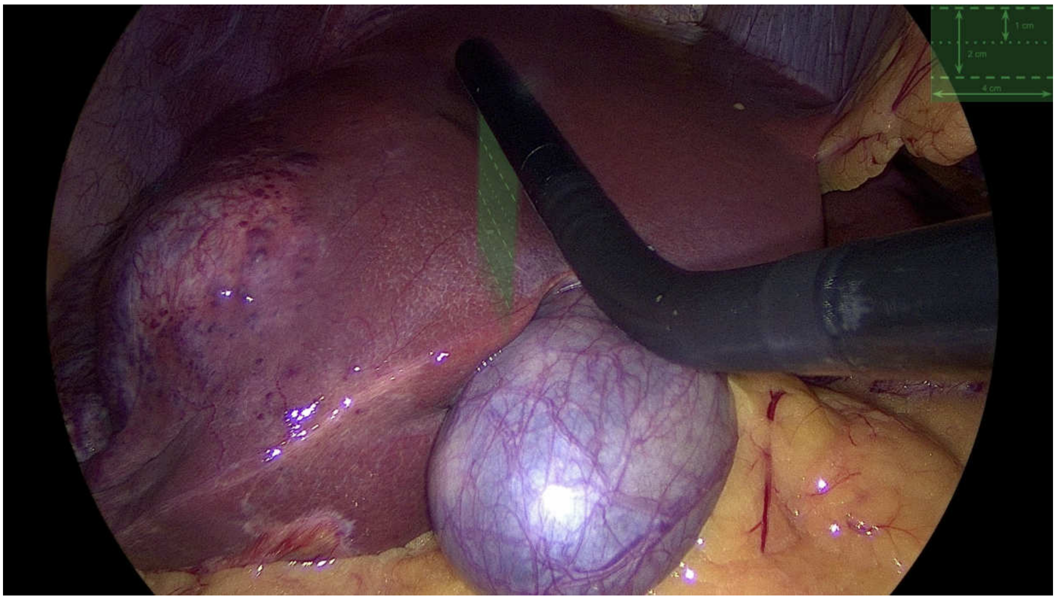


Figure 4.13: First visualization approach: displaying only LUS plane along with the scales.

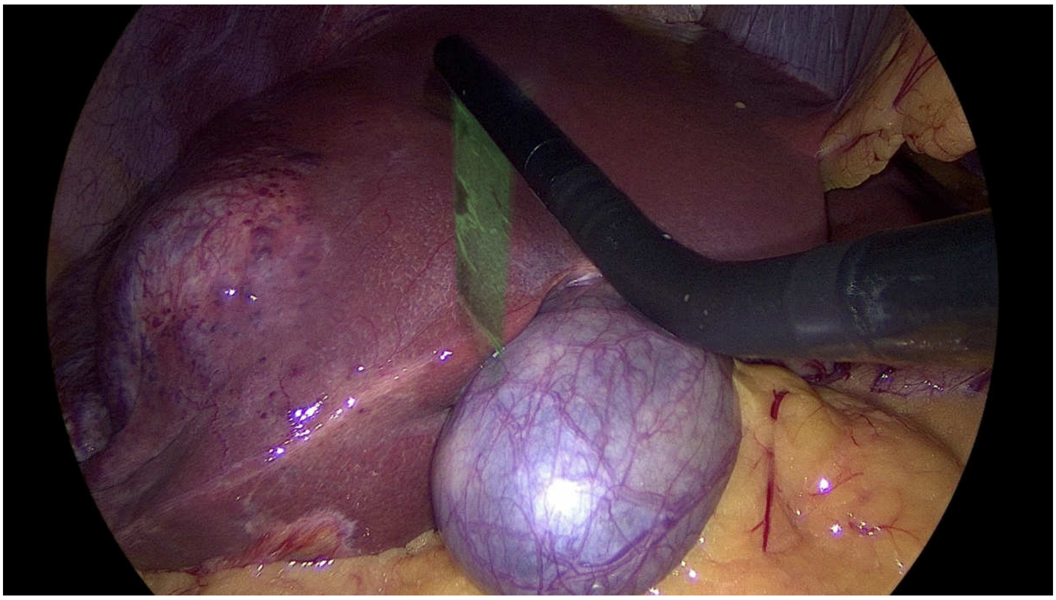


Figure 4.14: Second visualization approach: displaying LUS image by blending with the MIS image.

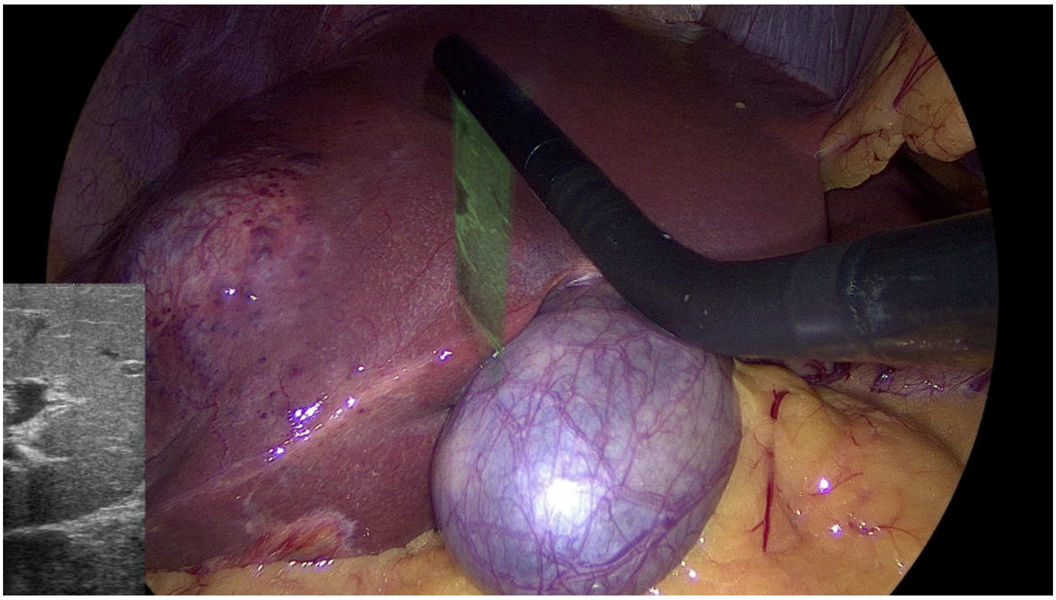


Figure 4.15: Third visualization approach: displaying the original LUS image on the corner of the blended images.

4.5 Conclusion

Given the breadth of topics addressed within this chapter, the concluding discussion is organized into thematic subsections to ensure a coherent and comprehensive synthesis of the findings.

4.5.1 Sensitivity to Noise

Depth-related estimates, such as the quadruplet $\alpha, \gamma, \theta, \|\mathbf{t}\|$, are approximately one order of magnitude more sensitive to image contour segmentation errors than lateral estimates. This heightened sensitivity arises from the typically large ratio between the distance of the LUS probe head to the laparoscope (approximately 10cm) and the LUS probe radius (approximately 0.5cm). This ratio effectively acts as an amplification factor for depth errors, which in turn leads to noticeable flickering in the augmented visualizations.

Consequently, improved image contour segmentation directly translates into more accurate depth-related estimates. In the absence of a more robust segmentation approach, we mitigate this effect by filtering the depth-related variables that are as decoupled from one another as possible. This issue is exacerbated in the `MSPDR` method, where the system state is not directly observable, and the estimation relies on sequences of noisy measurements.

4.5.2 Limitations

When both assumptions of `LARLUS` and `MSPDR` cannot be satisfied to resolve the 6th pose-DoF, the ensuing solutions can be mitigated by one or several of the following leads:

- *Solutions with user notification.* When the covariance associated with the missing sixth-DoF exceeds a predefined upper bound, the AR system can notify the surgeon, via textual cues, audio signals, or graphical overlays, to expose the LUS probe shaft in the camera field of view. This intervention enables the covariance to decrease below the threshold, allowing the sixth-DoF to be re-estimated with acceptable accuracy.

- *Solutions with robotic surgery.* One approach is to visually servo the robotic system to actively control the camera such that the LUS probe shaft remains within the field of view. Alternatively, if the LUS probe is also manipulated by a robot, the forward kinematics of both robots can be exploited, together with the known relative transformation between their base coordinate frames, to recover the full probe pose.

- *Solutions from computer vision.* One possibility is to estimate the LUS probe keyhole position on the patient's abdomen [Espinel *et al.* 2024]. Given the 5-DoF head pose and the known head length, the sixth-DoF can then be recovered by computing the intersection point between the head and shaft, followed by estimation of the shaft direction from this intersection point and the keyhole position. A second approach is to register a preoperative 3D organ model to the intraoperative LUS image [Montaña-Brown *et al.* 2021]. Under the assumption that the probe head is in contact with the organ surface, this registration provides an additional geometric constraint that can be used to resolve the missing pose-DoF. This idea is studied more in depth in the ensuing chapter.

4.5.3 Visualization

During discussions with the surgeons, several potential usage scenarios were identified, with individual preferences varying according to surgical workflow, personal experience, and the specific intraoperative context. While no single visualization mode was universally preferred, a consistent conclusion emerged from these discussions: flexibility in visualization is essential for practical clinical adoption. To address this, the proposed interface was designed to allow rapid switching between visualization modes, enabling the displayed information to be dynamically adapted to the surgeon's immediate preference and task requirements. This flexibility allows the system to accommodate different user expectations and procedural contexts without interrupting the surgical workflow, thereby improving usability and supporting more intuitive intraoperative interaction.

Chapter 5

Multimodal Registration and Full-tumor Augmentation

Contents

5.1 Introduction

MIS has significantly advanced liver surgery; however, accurate intraoperative tumor localization remains challenging. In LLR, this limitation is commonly addressed by incorporating two complementary modalities in addition to standard MIS imaging.

The first modality is LUS, which is regarded as the clinical gold standard. Despite its widespread use, LUS presents several limitations, including a steep learning curve, strong operator dependence, a limited field of view, and suboptimal image quality.

The second modality consists of preoperative three-dimensional anatomical models reconstructed from CT or MRI. Their intraoperative use is complicated by liver deformation caused by pneumoperitoneum and surgical manipulation. To mitigate these issues, AR approaches have recently been introduced [Labrunie *et al.* 2024, Espinel *et al.* 2024, Mhiri *et al.* 2024, Kalantari *et al.* 2024a, Ali *et al.* 2025]. However, existing methods typically rely on only two of the three available modalities, which leads to inherently ill-posed registration problems.

In this work, we propose a unified registration framework that simultaneously integrates all three modalities without relying on markers or external tracking systems, thereby maintaining OR sterility and workflow integrity. The proposed approach leverages novel constraints established between each pair of modalities to jointly estimate the deformation of the preoperative 3D model and the 6-DoF pose of the LUS probe. Experimental results on both phantom and patient datasets demonstrate that the proposed method outperforms state-of-the-art approaches based on only two modalities, as evidenced by qualitative and quantitative evaluations.

5.2 Related Work

We first review markerless registration approaches involving all pairwise combinations of two modalities, followed by methods that integrate all three modalities.

For the registration of preoperative 3D models to MIS images, early work [Robu *et al.* 2018] focused on rigid registration, which was subsequently shown to be insufficiently accurate for clinical use. This limitation motivated the development of deformable registration techniques [Mhiri *et al.* 2024, Labrunie *et al.* 2024, Espinel *et al.* 2024, Labrunie *et al.* 2022, Espinel *et al.* 2020, Adagolodjo *et al.* 2017, Koo *et al.* 2017]. These methods exploit correspondences between anatomical landmarks to guide the deformation process, resulting in substantially improved registration accuracy.

In preoperative 3D model to LUS registration, existing approaches primarily rely on vascular structures [Montaña-Brown *et al.* 2021, Ramalhinho *et al.* 2022] or incorporate kinematic priors [Ramalhinho *et al.* 2019]. Nevertheless, even when setting aside the 39% failure rate reported in the most recent study [Ramalhinho *et al.* 2022], such registration strategies remain inherently unstable due to intraoperative liver deformation.

For the registration of LUS to MIS images, prior approaches have relied on EM tracking [Liu *et al.* 2021], magneto-optic tracking [Feuerstein *et al.* 2007], or visual fiducial markers [Rabbani *et al.* 2022a]. More recently, markerless, vision-based registration methods have been introduced [Kalantari *et al.* 2024b, Kalantari *et al.* 2024a], alleviating the need for external tracking systems or artificial markers. Finally, two prior works have investigated the joint registration of all three modalities and are therefore the most closely related to our approach. The method introduced in [Montaña-Brown *et al.* 2022] is learning-based and patient-specific rigid registration, requiring retraining for each individual case and relying on synthetically rendered data. By contrast, our approach is patient-generic deformable registration and is formulated within a principled mathematical framework that explicitly models the relationships between modalities. This model-based design improves interpretability and facilitates step-wise error monitoring and control. Furthermore, the proposed method remains applicable even when the tumor is not visible in the LUS images. The approach presented in [Song *et al.* 2015] relies on EM tracking of the LUS probe, manual initialization, and vascular landmarks. However, EM tracking can be impractical in a surgical setting due to sterility constraints and the additional hardware clutter introduced into the operating room. In contrast, our method eliminates the need for manual initialization and operates exclusively with standard OR equipment, namely, the laparoscope and the LUS probe, thereby enabling seamless clinical integration.

5.3 Methodology

This section commences by modeling the problem and delineating the underlying assumptions. Subsequently, it transitions to a formal problem statement and the derivation of the cost functions, concluding with the proposal of minimization criteria for the cost function ensemble.

5.3.1 Modeling and Main Assumptions

We adopt the notations `Lap`, `Prp`, and `LUS` as concise shorthands to refer to the three considered modalities throughout this work. An overview of the principal inputs provided to the system, as well as the corresponding outputs it generates, is illustrated in Fig. 5.1. In our formulation, we assume that `LUS` and `Lap` are temporally synchronized; under this assumption, their relative spatial relationship can be accurately modeled by a rigid transformation that remains constant over time.

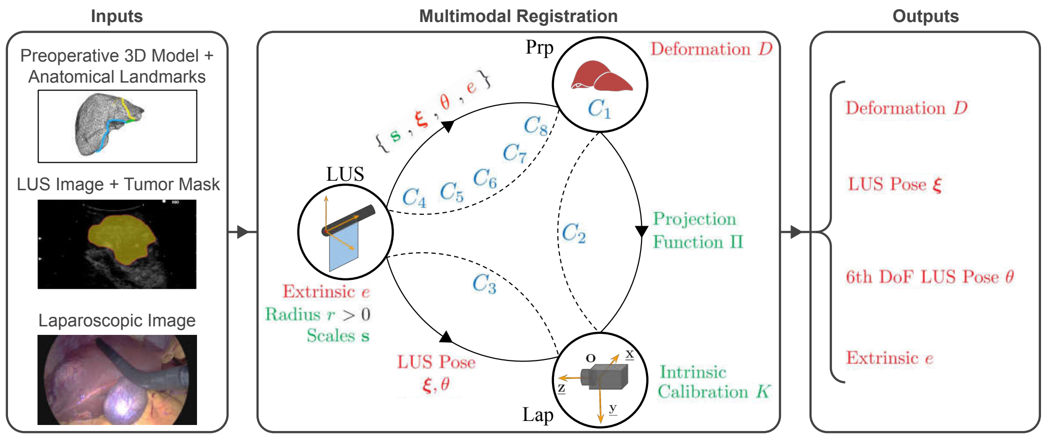


Figure 5.1: Modalities, parameters, and relationships. Optimizable parameters are in red, fixed ones in green, and cost terms for each pair of modalities (except C_1) in blue.

The `Lap` modality corresponds to the MIS camera, commonly referred to as the laparoscope. This modality contains an RGB image along with liver annotations. In our work, these annotations are obtained using the automatic segmentor [Labrunie *et al.* 2022], which identifies the liver silhouette, the falciform ligament, and the lower ridge. For the imaging model, we employ the standard pinhole camera formulation, where the intrinsic parameters are encapsulated in the 3×3 matrix K. In this matrix, f_x and f_y denote the focal lengths, while x_0 and y_0 correspond to the coordinates of the principal point in pixels. The camera is fully calibrated, and the captured images are undistorted using OpenCV routines. Consequently, the projection function can be expressed as $\Pi(X, Y, Z) = (f_x X/Z + x_0, f_y Y/Z + y_0)$.

For the purposes of registration, we adopt the standard pinhole coordinate frame as the reference frame, with its origin located at the camera center. *Under this convention, the registration process consists of determining the transformations necessary to map all other entities into the* `Lap` *coordinate frame*, thereby establishing a consistent spatial reference for subsequent analyses.

The `LUS` modality corresponds to the laparoscopic ultrasound probe. An ultrasound image is acquired when the transducer is in direct contact with the liver surface, yielding a gray-level image of the underlying tissue. In addition, a binary segmentation mask of the tumor is produced manually for each acquisition. In accordance with Fig. 2.7 and Fig. 3.1, the physical geometry of the probe is represented using a spherocylindrical model, consisting of a cylindrical section corresponding to the probe head and a spherical element modeling the probe tip, both characterized by a known radius $r > 0$.

The probe coordinate frame is defined with its origin located at the center of the spherical tip. The Z axis is aligned with the longitudinal axis of the cylindrical part of the probe, while the X axis is defined such that it intersects the linear transducer. Due to the inherent rotational symmetry of the spherocylindrical model about the Z axis, the full probe pose $(R \in SO_3, T \in \mathbb{R}^3)$ cannot be uniquely recovered from the laparoscopy image alone, which is a well-documented limitation in Chapter 3. To explicitly account for this ambiguity, the rotational component of the pose is parameterized using Euler angles and decomposed as $R_X(\alpha)R_Y(\beta)R_Z(\theta)$, allowing the pose to be represented in two distinct parts.

The first part comprises the 5-DoF that are observable from the laparoscopy image alone, denoted by $\boldsymbol{\xi} = (\alpha, \beta, T)$. The second part contains the remaining sixth degree of freedom, namely the rotation angle θ about the Z axis, which cannot be inferred without introducing an additional constraint. Furthermore, the probe model includes a single extrinsic parameter e, which specifies the offset of the ultrasound imaging plane along the Z axis. Finally, the intrinsic parameters of the ultrasound image are given by $\mathbf{s} = [s_x, s_y]$, representing the pixel-to-millimeter scaling factors that are directly read from the ultrasound display.

The `Prp` modality refers to the preoperative imaging data, which typically consist of either CT or MRI acquisitions. From these data, we assume that a three-dimensional preoperative model is reconstructed using MITK and represented as a triangulated mesh, whose vertices are denoted by $D_{\texttt{Prp}}$ and expressed in the preoperative coordinate frame. This model includes multiple anatomical components, namely the liver surface $\texttt{surf}(D_{\texttt{Prp}})$, the internal tumor structures, and salient anatomical landmarks such as the falciform ligament and the lower ridge.

The objective of the registration process is to determine an appropriate deformation of this preoperative mesh, resulting in a new set of vertices D expressed in the

Lap coordinate frame. Through this transformation, the preoperative anatomical model is spatially aligned with the intraoperative laparoscopic reference, enabling consistent integration of preoperative and intraoperative information.

5.3.2 Problem Statement and Cost Functions

Based on the modeling assumptions and representations introduced above, we formulate the registration task as a constrained optimization problem. Specifically, we define a global cost function C composed of eight individual terms C_i, each associated with a positive weighting factor $\lambda_i > 0$, for $i \in [1, 8]$, which controls the relative influence of the corresponding constraint.

The first term of the cost function represents a biomechanical energy that penalizes non-physiological deformations of the mesh and enforces plausible tissue behavior. The remaining terms encode constraints arising from pairwise interactions between the different modalities involved in the registration process. Among these, the first three terms are directly adapted from previously published work, whereas the last five terms constitute novel contributions introduced in this study. It is worth noting that certain cost terms may partially overlap in their effects, as some of them can be employed either during the initialization stage or during subsequent refinement, depending on the specific phase of the registration procedure.

Deformation energy (Prp) [Labrunie *et al.* 2024]. The cost term $C_1(D)$ is designed to preserve the structural and biomechanical coherence of the preoperative three-dimensional model throughout the registration process. To this end, it relies on a neo-Hookean material formulation, which penalizes unrealistic or non-physiological deformations and promotes mechanically plausible transformations of the mesh as it is deformed from its preoperative configuration.

Anatomical landmarks (Prp–Lap) [Labrunie *et al.* 2022]. The cost term $C_2(D)$ quantifies the discrepancy between the anatomical landmarks of the preoperative three-dimensional model and their corresponding features observed in the surgical image. More precisely, after the preoperative model is deformed and subsequently projected into the image plane, this term measures the distance between the projected landmark locations and their manually or automatically identified counterparts in the Lap image, thereby enforcing anatomical consistency across the two modalities.

Probe reprojection (Lap–LUS) [See Chapter 3] The cost term $C_3(\boldsymbol{\xi}) = \sum_{q \in \mathcal{S}} d(q, \Pi(\boldsymbol{\xi}))$ evaluates the consistency between the observed laparoscopic ultrasound probe contour and its geometric model. Specifically, it computes the cumulative point-to-model distance $d(\cdot, \cdot)$ between each point q belonging to the observed two-dimensional LUS contour $\mathcal{S}$ and the reprojection $\Pi(\boldsymbol{\xi})$ of the spherocylindrical probe model, parameterized by $\boldsymbol{\xi}$, into the surgical image.

Probe contact (Prp–LUS) - non-collision. The cost term $C_4(D, \boldsymbol{\xi})$ is introduced to explicitly enforce a non-collision constraint between the ultrasound probe and the liver, thereby ensuring that the probe geometry remains entirely outside the liver volume. To this end, we uniformly sample $N_{\texttt{probe}} = 520$ points over the surface of the LUS probe, which are collected in the set $\mathcal{P}_{\texttt{probe}}$.

For each sampled point $p \in \mathcal{P}\texttt{probe}$, the constraint relies on the signed distance function $d\pm(p, \texttt{surf}(D))$ computed with respect to the deformed liver surface $\texttt{surf}(D)$. By convention, the signed distance $d_{\pm}$ takes positive values when the point lies outside the liver and negative values when it penetrates the liver surface. The cost contribution associated with a given point is activated only when a penetration occurs, which is achieved by applying the Heaviside function $H(\cdot)$ to the negated signed distance. This formulation results in the following expression:

$$C_4(D, \boldsymbol{\xi}) = \sum_{p \in \mathcal{P}\texttt{probe}} H\left(-d \pm (p, \texttt{surf}(D))\right). \tag{5.1}$$

Overall, this term penalizes configurations in which any portion of the probe intersects the liver, thus promoting physically feasible probe–organ interactions during registration.

Probe contact (Prp–LUS) - position. The cost term $C_5(D, \boldsymbol{\xi})$ complements the non-collision constraint by explicitly modeling the positional relationship between the LUS probe and the liver surface. Rather than enforcing a strict exclusion, this term measures how close the probe lies to the organ, thereby encouraging physically consistent contact configurations. More specifically, the cost is defined as the cumulative distance between the sampled points of the probe and the deformed liver surface. Using the same set of probe samples $\mathcal{P}\texttt{probe}$, each point $p \in \mathcal{P}\texttt{probe}$ contributes through its distance $d(p, \texttt{surf}(D))$ to the liver surface $\texttt{surf}(D)$. This yields the following formulation:

$$C_5(D, \boldsymbol{\xi}) = \sum_{p \in \mathcal{P}_{\text{probe}}} d(p, \texttt{surf}(D)). \tag{5.2}$$

By minimizing this term, the optimization favors probe poses that are spatially consistent with contact or near-contact with the liver, while remaining compatible with the non-collision constraint enforced by C_4.

Probe contact (Prp–LUS) - orientation. The cost term $C_6(D, \boldsymbol{\xi}, \theta)$ encodes an orientation consistency constraint between the LUS probe and the liver surface at the region of contact. Specifically, it measures the absolute angular deviation $\text{ang}(\cdot, \cdot)$ between the imaging plane of the ultrasound probe and the local normals of the liver surface along the transducer, thereby promoting a physically plausible probe orientation during image acquisition. To compute this term, we sample $N_{\texttt{trans}} = 32$ points along the LUS transducer, which are stored in the set $\mathcal{P}\texttt{trans}$. For each sampled point $p \in \mathcal{P}\texttt{trans}$, the corresponding liver surface normal $\mathcal{N}(p, \texttt{surf}(D))$ is evaluated at the closest point on the deformed liver surface $\texttt{surf}(D)$. This normal is then compared to the probe orientation vector $\mathcal{U}(\boldsymbol{\xi}, \theta)$, which depends on both the recoverable pose parameters $\boldsymbol{\xi}$ and the additional rotational degree of freedom θ. The resulting cost is defined as:

$$C_6(D, \boldsymbol{\xi}, \theta) = \sum_{p \in \mathcal{P}_{\text{trans}}} \text{ang}\left(\mathcal{N}(p, \texttt{surf}(D)), \mathcal{U}(\boldsymbol{\xi}, \theta)\right). \tag{5.3}$$

Minimization of this term encourages alignment between the ultrasound imaging plane and the local liver surface geometry, which is consistent with typical probe–tissue interaction during laparoscopic ultrasound acquisition.

Tumor profile (`Prp-LUS`) - shape. The cost term $C_7(D, \boldsymbol{\xi}, \theta, e)$ enforces consistency between the tumor shape observed in laparoscopic ultrasound and the corresponding structure derived from the preoperative model. More precisely, it measures the discrepancy between the tumor profile extracted from the LUS image, denoted by $\mathcal{P}_{\texttt{prof}}$, and the two-dimensional slicing $U(D, \boldsymbol{\xi}, \theta, e)$ of the deformed preoperative three-dimensional model along the LUS imaging plane. To evaluate this term, we uniformly sample $N_{\texttt{prof}}$ points from the slicing U with a fixed spacing of 1mm, which typically yields $N_{\texttt{prof}} \in [30, 250]$ points depending on the tumor extent. The transformation matrix E is constructed from the LUS parameters $\boldsymbol{\xi}$, θ, and e, and maps coordinates from the LUS image frame to the base coordinate frame. In addition, the scale matrix $S = \text{Diag}(s_x, s_y)$ is assembled from the intrinsic scale vector $\mathbf{s}$. Using these transformations, a tumor profile point p from the LUS image is mapped into the `Lap` coordinate frame as $SE^{-1}p$.

The resulting cost is defined as:

$$C_7(D, \boldsymbol{\xi}, \theta, e) = \frac{1}{N_{\texttt{prof}}} \sum_{p \in \mathcal{P}_{\texttt{prof}}} \min_i \left(d(SE^{-1}(p - \bar{p}), U_i - \bar{u}) \right), \tag{5.4}$$

where the centroids $\bar{p} = \text{mean}(\mathcal{P})$ and $\bar{u} = \text{mean}(U)$ are subtracted to ensure translation invariance. This formulation encourages agreement between the LUS-derived tumor contour and the corresponding preoperative tumor slice, thereby aligning tumor shape information across modalities.

Tumor profile (`Prp-LUS`) - position. The cost term $C_8(D, \boldsymbol{\xi}, \theta, e) = \left| SE^{-1}\bar{p} - \bar{u} \right|$ captures the positional alignment between the tumor observed in laparoscopic ultrasound and its counterpart in the preoperative model. Specifically, it computes the Euclidean distance between the centroid $\bar{p}$ of the LUS-derived tumor profile and the centroid $\bar{u}$ of the corresponding slice from the deformed preoperative 3D model, after transforming the LUS centroid into the `Lap` coordinate frame using the scale and transformation matrices S and E. By minimizing this term, the registration ensures that the overall tumor location is consistent across modalities, complementing the shape alignment enforced by C_7.

5.3.3 Minimization Method

The optimization problem is characterized by a high-dimensional parameter space, typically comprising between 600 and 800 unknowns, and incorporates eight cost terms of heterogeneous nature. These terms exhibit varying properties, including smoothness and non-continuous differentiability, with the majority being non-convex. Such complexity makes random initialization insufficient for reliable convergence. To address this, we propose a structured initialization and refinement strategy, as detailed in Algorithm 5, which systematically guides the optimization towards physically and anatomically plausible solutions.

Algorithm 5: Optimization Criterion

Inputs: $D_{\texttt{Prp}}$ *(initial liver mesh),* $\mathbf{s}$ *(LUS scales),* I *(laparoscopic image),* K *(laparoscope intrinsics)*

Outputs: D *(deformed liver),* $\boldsymbol{\xi}$ *(5-DoF LUS pose),* θ *(6th-DoF LUS pose),* e *(LUS extrinsic)*

/* Initialization */

$D \leftarrow$ `reg_Prp_to_Lap`$(D_{\texttt{Prp}}, I, K)$ `// Use [Labrunie et al. 2022]`

$\boldsymbol{\xi} \leftarrow$ `estimate_5DoF_LUS_Pose`(I, K) `// See Chapter 3`

$D, \boldsymbol{\xi} \leftarrow \arg\min_{D,\boldsymbol{\xi}} C_4(D, \boldsymbol{\xi}) + C_5(D, \boldsymbol{\xi})$ `// Probe-liver contact`

$[\theta, e] \leftarrow [\arg\min_{\theta} C_6(D, \boldsymbol{\xi}, \theta), 0]$ `// Transducer-liver contact`

/* Refinement */

converged $\leftarrow$ **False**

while not converged **do**

    /* Differential evolution minimization */

    $\boldsymbol{\xi}, \theta, e \leftarrow \arg\min_{\boldsymbol{\xi},\theta,e} C_7(D, \boldsymbol{\xi}, \theta, e) + \lambda_3 C_3(\boldsymbol{\xi})$

    /* Gauss-Newton minimization */

    $D^- \leftarrow D$ `// Save current deformation`

    $D \leftarrow \arg\min_D C_1(D) + \lambda_2 C_2(D) + \lambda_4 C_4(D, \boldsymbol{\xi}) + \lambda_5 C_5(D, \boldsymbol{\xi}) + \lambda_8 C_8(D, \boldsymbol{\xi}, \theta, e)$

    if $\|D - D^-\| < \tau$ **then** converged $\leftarrow$ **True** `// Convergence verification`

end

Initialization. The initialization procedure leverages the inherent dependencies between variables as well as available solution methods to produce a physically plausible starting point for the optimization. The process begins by applying the method of [Labrunie *et al.* 2022] to obtain an initial estimate of the deformation parameters D for the preoperative model. This is followed by the approach of Chapter 3, which provides an initial estimate of the LUS probe's 5-DoF pose $\boldsymbol{\xi}$.

Subsequently, a first layer of constraints linking the `LUS` and `Prp` modalities is introduced through the cost terms C_4 and C_5, which enforce proper contact between the LUS probe and the liver surface. These constraints are used to iteratively update both D and $\boldsymbol{\xi}$, ensuring that the probe configuration is consistent with the organ geometry.

Finally, the remaining LUS parameters that are not directly observable are initialized while considering the entire set of optimization variables. In particular, the unobserved rotation angle θ of the probe is initialized based on the orientation of the liver surface normals to approximate correct alignment, and the extrinsic imaging plane offset e is set to zero, positioning the transducer near the probe tip. This structured initialization establishes a robust and anatomically coherent starting point for the subsequent refinement stage.

Refinement. The refinement stage decomposes the overall minimization into two sequential components based on the mathematical properties of the cost terms. In the first component, we focus on the non-smooth terms C_3 and C_7, which are optimized using a differential evolution algorithm capable of handling non-differentiable and potentially discontinuous objectives. In the second component, the remaining smooth terms, comprising all other cost contributions, are optimized using a Gauss-Newton method, which leverages gradient information to efficiently converge to a local minimum. Both optimization stages are implemented using the SciPy library.

To enhance computational efficiency and improve numerical stability, we employ dimensionality reduction techniques [Labrunie *et al.* 2024], limiting the number of mesh vertices considered in the optimization to 200. Additionally, we explicitly code the Jacobian matrix, ensuring accurate and rapid evaluation during the Gauss-Newton step. Convergence of the refinement process is declared when the change in deformation parameters falls below a predefined threshold τ, indicating that the solution has stabilized.

5.4 Experimental Results

All hyperparameters were fixed to constant values determined empirically. Specifically, the weighting factors for the cost terms were set as $\lambda_2 = 0.1$, $\lambda_3 = \lambda_4 = \lambda_5 = 10$, and $\lambda_8 = 3$. The convergence threshold for the refinement step was chosen as $\tau = 3.5$.

5.4.1 Phantom Study

In this study, we evaluated the proposed co-registration framework using eight distinct deformation states of a three-dimensional liver model containing a virtual tumor. These deformations were generated following the methodology of [Ribeiro *et al.* 2024], as introduced in Sec. 2.2.2, in order to simulate realistic intraoperative organ shape variations. Each deformed liver configuration was subsequently materialized through 3D printing, yielding a set of physical phantoms with known geometry that served as controlled experimental surrogates. This design provided precise GT information for both tumor location and anatomical landmarks, as described in Sec. 2.3.4, thereby enabling quantitative evaluation under repeatable and anatomically consistent conditions. An example of the experimental setup is shown in Fig. 5.2a.

To emulate intraoperative ultrasound acquisition, a real LUS probe was used during the experiments. The initial pose of the probe was estimated using an ArUco marker-based initialization, providing a reliable reference for subsequent simulation steps. The corresponding US imaging plane was then intersected with the GT tumor volume to generate a simulated tumor contour, to which Gaussian noise was added in order to approximate realistic sensing uncertainty and segmentation variability. Anatomical landmarks visible in the laparoscopic images were manually annotated, and the remaining stages of the experiment were conducted according

	# Phantom								
	1	2	3	4	5	6	7	8	Avg
Base	37.68	23.44	23.46	17.01	20.92	21.08	20.52	23.64	23.47
+ Contact	-26%	-3%	-6%	+23%	+9%	+3%	-18%	-15%	-4%
+ All	-49%	-72%	-36%	-55%	-35%	-48%	-24%	-22%	-43%

Table 5.1: TRE (mm) for the base method [Labrunie *et al.* 2022] in the phantom study, with percentage changes from contact-only and all proposed constraints.

to the procedure described in Algorithm 5. It is important to emphasize that the simulated LUS measurements correspond to the tumor position in the deformed liver state rather than the original undeformed GT configuration, thereby reflecting the clinically relevant registration scenario.

The quantitative registration results are summarized in Table 5.1. The estimated LUS probe rotation angle θ achieved an Mean Absolute Error (MAE) of 10.3° with a standard deviation of 6.3°, while the estimated extrinsic parameter e yielded an MAE of 1.2mm with a standard deviation of 0.4mm. Most importantly, the integration of laparoscopic imagery, LUS, and preoperative model information within the proposed co-registration framework resulted in a substantial improvement in tumor localization accuracy. Specifically, the method achieved an average reduction of 43% in TRE relative to existing approaches, demonstrating the effectiveness of multimodal integration for improving intraoperative tumor localization under organ deformation.

5.4.2 Public Benchmark

We further evaluated the proposed method on the publicly available dataset introduced in Sec. 2.2.1, originally published in. [Rabbani *et al.* 2022a]. This dataset enables direct comparison of tumor localization accuracy across competing methods under a common evaluation protocol and additionally provides GT information for the LUS imaging plane through fiducial markers attached to the probe. Importantly, although these markers are available for evaluation and GT generation, they are not used by the proposed method at inference time. An example of the experimental setup is shown in Fig. 5.2b.

The quantitative results for three patient cases are summarized in Table 5.2. For patient P1, the proposed method yields a modest but consistent improvement over the previously best reported result [Adagolodjo *et al.* 2017], reducing the tumor TRE from 8.25mm to 7.76mm. For patient P3, the improvement is considerably more pronounced, with the tumor TRE decreasing from 17.60mm, as reported in [Labrunie *et al.* 2024], to 8.58mm, corresponding to a substantial gain in localization accuracy. For patient P4, however, the proposed method underperforms relative to the best previously reported result [Labrunie *et al.* 2022], with the tumor TRE increasing from 7.23mm to 12.35mm.

Overall, the proposed framework demonstrates strong performance when the intraoperative tumor morphology remains consistent with the preoperative model. Excluding patient P4, for whom a substantial discrepancy between the preoperative and intraoperative tumor appearance was observed, the average tumor TRE across cases is reduced by 53% relative to the best prior reported performance [Koo *et al.* 2017], decreasing from 17.27mm to 8.17mm. These results indicate that the proposed method is particularly effective when the preoperative model provides a reliable approximation of the intraoperative anatomy. Moreover, the framework remains applicable in scenarios involving moderate tumor progression or morphological change between preoperative imaging and surgery, although such deviations may reduce registration accuracy. The estimated LUS probe rotation angle θ achieved a mean absolute error of 8.7° with a standard deviation of 4.7°, further confirming the stability of the proposed pose refinement strategy.

5.4.3 Qualitative Study

We performed a qualitative evaluation on two additional patients, using a total of five images, as illustrated in Fig. 5.2c. The study was structured around four validation steps.

(i) The 5-DoF LUS pose $\boldsymbol{\xi}$ was assessed by comparing the silhouette of the reprojected LUS probe model, using the optimized pose parameters, with the segmented probe in the `Lap` image. Across all experiments, the average Intersection over Union (IoU) achieved was 97%, confirming accurate pose estimation.

(ii) The 6th-DoF, corresponding to the LUS rotation angle θ, was evaluated by uniformly sampling candidate values over $[0, 2\pi)$ in increments of 0.1°. For each sampled angle, the distance between the tumor observed in the LUS image

	Tumor TRE (mm)						
	MA	M1	M2	M3	LMR	NM	Ours
P1	15.14	8.25	9.49	14.87	17.40	14.82	**7.76**
P3	30.48	28.40	25.04	22.40	17.60	20.15	**8.58**
P4 †	16.29	15.83	18.35	**7.23**	17.00	12.95	12.35 †
Avg. ~~P4~~	22.81	18.33	17.27	18.64	17.50	17.49	**8.17**
Avg.	20.63	17.49	17.63	14.83	17.33	15.97	**9.56**

† P4 has gone through “tumor evolution”

Table 5.2: Tumor TRE on public benchmark [Rabbani *et al.* 2022a]. “Tumor evolution” indicates a change in tumor size between preoperative imaging and surgery. MA is manual rigid registration, M1 is [Adagolodjo *et al.* 2017], M2 is [Koo *et al.* 2017], M3 is [Labrunie *et al.* 2022], LMR is [Labrunie *et al.* 2024], and NM is [Mhiri *et al.* 2024].

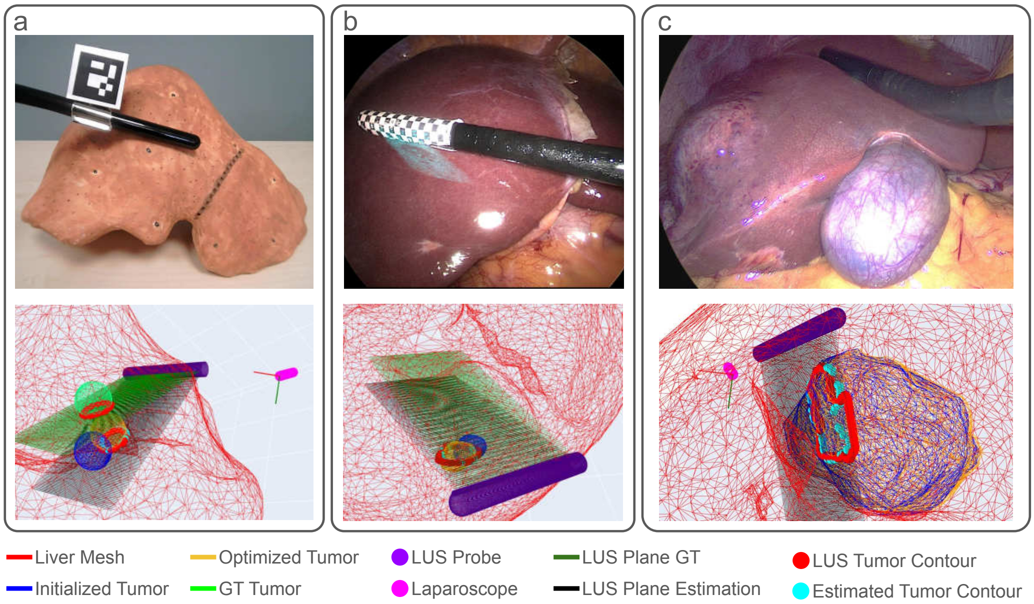


Figure 5.2: Experiments: (a) phantom, (b) public benchmark, (c) patients qualitative.

and the corresponding tumor cross-section from the preoperative model (`Prp`) was computed. The optimized θ consistently produced the minimal distance, with an average of 1.2mm, demonstrating its near-optimality.

(iii) Liver deformation accuracy was quantified by measuring the distance between anatomical landmarks reprojected from the optimized model and their corresponding locations in the `Lap` images. The mean distance was 31 ± 6 pixels, which is reasonable given the high resolution of the laparoscopic images, indicating plausible deformation modeling.

(iv) Tumor shape and translation were evaluated both before and after optimization. Significant deformation, defined as a volume change exceeding 20%, would indicate inconsistencies in the optimization process. Similarly, insufficient tumor translation relative to the desired contour would suggest suboptimal initialization, requiring local adjustments constrained by other optimization parameters. In our experiments, no substantial tumor deformation or inadequate translation was observed, confirming the robustness and consistency of the proposed optimization framework.

5.4.4 Discussion on the Experiments

Validating registration during surgery remains a highly challenging task. Our method has been evaluated retrospectively on a cohort of six patients and eight phantoms. Qualitative results from two patients, together with both qualitative and quantitative results for the eight phantoms and Patients P1 and P3 from [Rabbani *et al.* 2022a], demonstrate the advantages of integrating all three standard modalities in the registration process.

As discussed in Sec. 5.4.2, Patient P4 illustrates the method's sensitivity to the similarity between the preoperative tumor model and the actual tumor present on the surgery day. When the preoperative model diverges significantly from intraoperative reality, the method continues to operate but introduces some registration errors. This behavior reflects an intrinsic limitation: if a substantial mismatch exists between preoperative and intraoperative models, augmented reality guidance should not be applied at all.

Furthermore, Patient P2, excluded from Table 5.2, highlights the critical role of the underlying deformation model. In cases where the base model [Labrunie *et al.* 2022] fails to converge during initialization, subsequent optimization becomes infeasible, emphasizing the dependency on a robust initial model.

From a computational perspective, the optimization algorithm completes in approximately 4 ± 1 minutes on a standard laptop. When combined with automated intraoperative image annotation [Labrunie 2024] and manual tumor annotation in the LUS image, the full registration workflow can be completed in under six minutes. This runtime confirms the clinical feasibility of the method for single-shot augmented reality guidance immediately prior to liver resection.

5.5 Conclusion

This study presented a novel multimodal framework for tumor localization in minimally invasive liver surgery that leverages the complementary strengths of laparoscopic imaging, laparoscopic ultrasound, and preoperative anatomical models within a unified co-registration formulation. In contrast to conventional navigation strategies, the proposed approach operates without fiducial markers, external tracking systems, or additional intraoperative hardware, thereby preserving workflow simplicity while maintaining practical compatibility with standard surgical practice. By jointly integrating visual, ultrasonic, and model-based information, the method enables robust intraoperative tumor localization under anatomically challenging conditions, including organ deformation and partial observability.

Comprehensive validation on both controlled phantom experiments and real patient data demonstrated that the proposed framework achieves high registration accuracy and consistently outperforms existing state-of-the-art approaches in tumor localization performance. These quantitative gains were obtained while maintaining a clinically acceptable runtime, supporting the practical feasibility of the method for intraoperative use. Beyond numerical performance, the method also received positive qualitative feedback from surgeons with respect to interpretability and integration into the surgical workflow, further reinforcing its translational relevance. Taken together, these results highlight the strong potential of the proposed framework to improve intraoperative tumor localization and to serve as a clinically viable foundation for image-guided assistance in minimally invasive liver surgery.

Chapter 6

Conclusion and Future Research

This thesis has presented a complete framework for multimodal registration and augmented reality guidance in laparoscopic liver resection, with the objective of improving intraoperative perception of subsurface anatomy through the principled integration of laparoscopic ultrasound into the surgical workflow. The central premise of this work is that laparoscopic ultrasound can serve not only as a diagnostic adjunct, but as the missing geometric and anatomical bridge between preoperative imaging and the intraoperative laparoscopic scene. By leveraging this modality as a real-time source of subsurface information, this thesis addressed one of the core limitations of minimally invasive liver surgery: the inability to directly perceive hidden tumors and critical vascular structures during resection.

To achieve this objective, the thesis has introduced and validated three complementary methodological contributions. First, a marker-free and sensor-free framework has been proposed for estimating the 5-DoF pose of laparoscopic ultrasound probes directly from monocular laparoscopic video. This contribution addressed a practical and longstanding limitation in the field by eliminating the need for external trackers, fiducials, or cumbersome intraoperative instrumentation, while maintaining real-time compatibility and intraoperative usability. Second, the underconstrained nature of monocular spherocylindrical pose recovery has been addressed through two complementary strategies for disambiguating and stabilizing the sixth degree of freedom. By combining geometric constraints with motion priors and temporal filtering, the resulting framework enabled robust 6-DoF tracking and stable in situ visualization of the ultrasound imaging plane. Third, this work has introduced a multimodal registration framework that incorporated ultrasound-derived anatomical observations and probe pose constraints to align preoperative CT/MRI models with intraoperative data, enabling coherent and spatially consistent full-tumor augmentation within the laparoscopic scene. Together, these contributions have established a complete pipeline for sensor-free and clinically compatible augmented reality guidance in laparoscopic liver surgery.

A major strength of this work lies in its methodological and experimental completeness. Beyond the proposed algorithms, this thesis has established the validation infrastructure required for rigorous assessment in a domain where ground truth is intrinsically difficult to obtain. To this end, it has introduced a diverse multimodal evaluation framework spanning clinical, ex vivo, semi-synthetic, and phantom datasets, each designed to capture complementary aspects of the problem. These resources enabled controlled benchmarking, uncertainty quantification, and progressive validation under increasing anatomical and procedural complexity. This experimental design ensured that the proposed methods were not only theoretically sound, but also robust under conditions representative of clinical practice.

Despite these advances, several challenges remain before such systems can be deployed routinely in clinical practice. The first concerns deformable modeling. While the proposed multimodal registration framework compensates for part of the discrepancy between preoperative and intraoperative anatomy through ultrasound-driven constraints, liver deformation remains only partially modeled and remains one of the principal limitations of current image-guided liver surgery. Future work should therefore investigate biomechanically informed and data-driven deformation models capable of integrating sparse intraoperative observations in real time. Such models would improve not only tumor localization, but also the spatial consistency of vascular and parenchymal augmentations throughout the procedure.

A second challenge concerns robustness and generalization in clinical environments. Although the proposed vision-based framework has been validated across varied datasets, operating room conditions remain highly variable, with frequent occlusions, specularities, smoke, blood, motion blur, and tool clutter. Improving robustness under these conditions will require tighter integration between geometric reasoning and learning-based perception, particularly for segmentation, pose recovery, and uncertainty estimation. In this context, hybrid methods that combine analytical priors with learned feature representations constitute a promising direction for improving both robustness and interpretability.

A third avenue for future research lies in extending the proposed framework beyond passive guidance toward active and adaptive intraoperative assistance. In particular, the integration of this work with robotic laparoscopic ultrasound manipulation, as envisioned in the broader IMMORTALLS project, would enable closed-loop tumor tracking and autonomous maintenance of optimal ultrasound views. Such a system would reduce cognitive load, improve consistency of intraoperative imaging, and move surgical guidance from visualization toward active assistance. More broadly, coupling multimodal registration with robotic perception and control opens a compelling path toward intelligent intraoperative navigation systems.

Finally, future work should also focus on clinical translation and human factors. The ultimate value of augmented reality guidance is not determined solely by geometric accuracy, but by its ability to improve surgical decision-making without disrupting workflow. Prospective evaluation with surgeons, ergonomic studies, and task-based assessments will therefore be essential to determine how such systems influence attention, confidence, cognitive load, and operative performance. The transition from technical feasibility to clinical utility will depend as much on usability and trust as on algorithmic precision.

In conclusion, this thesis has demonstrated that laparoscopic ultrasound can serve as a practical and effective anchor for multimodal intraoperative guidance in laparoscopic liver surgery. By combining real-time vision-based probe tracking, ultrasound-aware multimodal registration, and augmented reality visualization into a unified framework, this work advances the state of the art toward clinically viable, sensor-free, and anatomically informed surgical navigation. Beyond its immediate contributions to laparoscopic liver resection, this work provides a generalizable foundation for multimodal guidance in minimally invasive surgery and contributes one step toward a broader paradigm of intelligent, perceptually augmented surgical systems.

Bibliography

[Abdel-Misih & Bloomston 2010] Sherif R.Z. Abdel-Misih and Mark Bloomston. *Liver Anatomy*. Surgical Clinics of North America, vol. 90, no. 4, page 643–653, August 2010. (Cited on page 3.)

[Abdulkareem *et al.* 2023] Nashwan Karkhi Abdulkareem, Shereen Ismail Hajee, Fatiheea Fatihalla Hassan, Ilham Khalid Ibrahim, Ruaa Emad Hussein Al-Khalidi and Noor Abubaker Abdulqader. *Investigating the slice thickness effect on noise and diagnostic content of single-source multi-slice computerized axial tomography*. Journal of Medicine and Life, vol. 16, no. 6, page 862–867, June 2023. (Cited on page 6.)

[Adagolodjo *et al.* 2017] Yinoussa Adagolodjo, Raffaella Trivisonne, Nazim Haouchine, Stéphane Cotin and Hadrien Courtecuisse. *Silhouette-based pose estimation for deformable organs application to surgical augmented reality*. pages 539–544, Sept 2017. (Cited on pages 70, 78 and 79.)

[Al Farai *et al.* 2024] Abdallah Al Farai, Mohammad Al Farei and Adil Al Zadjali. Perspective chapter: Open versus minimally invasive surgery accessibility in different economic level countries. IntechOpen, November 2024. (Cited on page 2.)

[Ali *et al.* 2025] Sharib Ali, Yamid Espinel, Yueming Jin, Peng Liu, Bianca Güttner, Xukun Zhang, Lihua Zhang, Tom Dowrick, Matthew J. Clarkson, Shiting Xiao, Yifan Wu, Yijun Yang, Lei Zhu, Dai Sun, Lan Li, Micha Pfeiffer, Shahid Farid, Lena Maier-Hein, Emmanuel Buc and Adrien Bartoli. *An objective comparison of methods for augmented reality in laparoscopic liver resection by preoperative-to-intraoperative image fusion from the MICCAI2022 challenge*. Medical Image Analysis, vol. 99, page 103371, 2025. (Cited on page 69.)

[Allan *et al.* 2015] Max Allan, Ping-Lin Chang, Sébastien Ourselin, David J. Hawkes, Ashwin Sridhar, John Kelly and Danail Stoyanov. *Image Based Surgical Instrument Pose Estimation with Multi-class Labelling and Optical Flow*. In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, pages 331–338. Springer International Publishing, 2015. (Cited on pages 28, 29 and 31.)

[Allan *et al.* 2018] M. Allan, S. Ourselin, D. J. Hawkes, J. D. Kelly and D. Stoyanov. *3-D Pose Estimation of Articulated Instruments in Robotic Minimally Invasive Surgery*. IEEE Transactions on Medical Imaging, vol. 37, no. 5, pages 1204–1213, 2018. (Cited on pages 29 and 31.)

[Alleblas *et al.* 2017] Chantal C. J. Alleblas, Michel P. H. Vleugels, Sjors F. P. J. Coppus and Theodoor E. Nieboer. *The effects of laparoscopic graspers with enhanced haptic feedback on applied forces: a randomized comparison with conventional graspers.* Surgical Endoscopy, vol. 31, no. 12, page 5411–5417, June 2017. (Cited on page 3.)

[Barfoot 2024] Timothy D Barfoot. State estimation for robotics. Cambridge University Press, Cambridge, England, 2 édition, February 2024. (Cited on page 53.)

[Birlo *et al.* 2022] Manuel Birlo, P.J. Eddie Edwards, Matthew Clarkson and Danail Stoyanov. *Utility of optical see-through head mounted displays in augmented reality-assisted surgery: A systematic review.* Medical Image Analysis, vol. 77, page 102361, April 2022. (Cited on page 8.)

[Bouget *et al.* 2017] David Bouget, Max Allan, Danail Stoyanov and Pierre Jannin. *Vision-based and marker-less surgical tool detection and tracking: a review of the literature.* Medical Image Analysis, vol. 35, pages 633–654, jan 2017. (Cited on page 28.)

[Buell *et al.* 2009] Joseph F. Buell, Daniel Cherqui, David A. Geller, Nicholas O'Rourke, David Iannitti, Ibrahim Dagher, Alan J. Koffron, Mark Thomas, Brice Gayet, Ho Seong Han, Go Wakabayashi, Giulio Belli, Hironori Kaneko, Chen-Guo Ker, Olivier Scatton, Alexis Laurent, Eddie K. Abdalla, Prosanto Chaudhury, Erik Dutson, Clark Gamblin, Michael D'Angelica, David Nagorney, Giuliano Testa, Daniel Labow, Derrik Manas, Ronnie T. Poon, Heidi Nelson, Robert Martin, Bryan Clary, Wright C. Pinson, John Martinie, Jean-Nicolas Vauthey, Robert Goldstein, Sasan Roayaie, David Barlet, Joseph Espat, Michael Abecassis, Myrddin Rees, Yuman Fong, Kelly M. McMasters, Christoph Broelsch, Ron Busuttil, Jacques Belghiti, Steven Strasberg and Ravi S. Chari. *The International Position on Laparoscopic Liver Surgery: The Louisville Statement, 2008.* Annals of Surgery, vol. 250, no. 3, page 825–830, November 2009. (Cited on page 3.)

[Carion *et al.* 2026] Nicolas Carion, Laura Gustafson, Yuan-Ting Hu, Shoubhik Debnath, Ronghang Hu, Didac Suris, Chaitanya Ryali, Kalyan Vasudev Alwala, Haitham Khedr, Andrew Huang, Jie Lei, Tengyu Ma, Baishan Guo, Arpit Kalla, Markus Marks, Joseph Greer, Meng Wang, Peize Sun, Roman Rädle, Triantafyllos Afouras, Effrosyni Mavroudi, Katherine Xu, Tsung-Han Wu, Yu Zhou, Liliane Momeni, Rishi Hazra, Shuangrui Ding, Sagar Vaze, Francois Porcher, Feng Li, Siyuan Li, Aishwarya Kamath, Ho Kei Cheng, Piotr Dollár, Nikhila Ravi, Kate Saenko, Pengchuan Zhang and Christoph Feichtenhofer. *SAM 3: Segment Anything with Concepts*, 2026. (Cited on pages 42 and 58.)

[Chandelon & Bartoli 2024] Kilian Chandelon and Adrien Bartoli. *Synchronising a stereoscopic surgical video stream using specular reflection.* International

Journal of Computer Assisted Radiology and Surgery, vol. 20, no. 2, page 289–299, July 2024. (Cited on page 5.)

[Chanwat *et al.* 2025] Rawisak Chanwat, Chairat Bunchaliew, Nuengruetai Olarnnapalai, Rapheephat Tanompetsanga, Sarat Sanguanlosit, Kittipong Chaiyabutr and Saard Treepongkaruna. *Sheep Liver as a Model for Advanced Laparoscopic Hepatic Resection Training*. Asian Journal of Endoscopic Surgery, vol. 18, no. 1, January 2025. (Cited on page 19.)

[Chaurasia & Culurciello 2017] Abhishek Chaurasia and Eugenio Culurciello. *LinkNet: Exploiting encoder representations for efficient semantic segmentation*. 2017 IEEE Visual Communications and Image Processing (VCIP), pages 1–4, 2017. (Cited on page 34.)

[Cherfioui 2025] Karima Cherfioui. *Optimization, development, and validation of vision-based ultrasound probe pose estimation pipeline*, 2025. Unpublished internal university report. (Cited on page 35.)

[Coelho 2016] Fabricio Ferreira Coelho. *Laparoscopic liver resection: Experience based guidelines*. World Journal of Gastrointestinal Surgery, vol. 8, no. 1, page 5, 2016. (Cited on page 3.)

[Cross & Zisserman 1998] G. Cross and A. Zisserman. *Quadric reconstruction from dual-space geometry*. In Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271), pages 25–31, 1998. (Cited on page 31.)

[Dellaert & Kaess 2017] Frank Dellaert and Michael Kaess. *Factor Graphs for Robot Perception*. Foundations and Trends in Robotics, vol. 6, no. 1-2, page 1–139, 2017. (Cited on pages 52 and 53.)

[Dirie *et al.* 2018] Najib Isse Dirie, Qing Wang and Shaogang Wang. *Two-Dimensional Versus Three-Dimensional Laparoscopic Systems in Urology: A Systematic Review and Meta-Analysis*. Journal of Endourology, vol. 32, no. 9, page 781–790, 2018. (Cited on page 4.)

[Doignon & Mathelin 2007] C. Doignon and M. Mathelin. *A Degenerate Conic-Based Method for a Direct Fitting and 3-D Pose of Cylinders with a Single Perspective View*. IEEE International Conference on Robotics and Automation, pages 4220–4225, 2007. (Cited on pages 29 and 31.)

[Dzubiella *et al.* 2000] Joachim Dzubiella, Matthias Schmidt and Hartmut Löwen. *Topological defects in nematic droplets of hard spherocylinders*. Physical Review E, vol. 62, no. 4, pages 5081–5091, October 2000. (Cited on page 28.)

[Espinel *et al.* 2020] Yamid Espinel, Erol Özgür, Lilian Calvet, Bertrand Le Roy, Emmanuel Buc and Adrien Bartoli. *Combining Visual Cues with Interactions* *for 3D–2D Registration in Liver Laparoscopy.* Annals of Biomedical Engineering, vol. 48, no. 6, page 1712–1727, February 2020. (Cited on pages 4 and 70.)

[Espinel *et al.* 2024] Yamid Espinel, Navid Rabbani, Thien Bao Bui, Mathieu Ribeiro, Emmanuel Buc and Adrien Bartoli. *Keyhole-aware laparoscopic augmented reality.* Medical Image Analysis, vol. 94, page 103161, 2024. (Cited on pages 67, 69 and 70.)

[Feuerstein *et al.* 2007] Marco Feuerstein, Tobias Reichl, Jakob Vogel, Armin Schneider, Hubertus Feussner and Nassir Navab. *Magneto-optic tracking of a flexible laparoscopic ultrasound transducer for laparoscope augmentation.* In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2007, pages 458–466. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. (Cited on pages 48 and 70.)

[Fitzgibbon *et al.* 1999] A. Fitzgibbon, M. Pilu and R.B. Fisher. *Direct least square fitting of ellipses.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, no. 5, pages 476–480, 1999. (Cited on page 31.)

[Fox 2025] Matthew Fox. *Is minimally invasive surgery at an inflection point?* ACS Bulletin, vol. 110, no. 9, October 2025. Accessed: 25 March 2026. (Cited on page 1.)

[Fox 2026] Matthew Fox. *Cost of robotic surgery remains complex equation.* ACS Bulletin, vol. 111, February 2026. Accessed: 01 April 2026. (Cited on page 2.)

[Fretland *et al.* 2018] Åsmund Avdem Fretland, Vegar Johansen Dagenborg, Gudrun Maria Waaler Bjørnelv, Airazat M. Kazaryan, Ronny Kristiansen, Morten Wang Fagerland, John Hausken, Tor Inge Tønnessen, Andreas Abildgaard, Leonid Barkhatov, Sheraz Yaqub, Bård I. Røsok, Bjørn Atle Bjørnbeth, Marit Helen Andersen, Kjersti Flatmark, Eline Aas and Bjørn Edwin. *Laparoscopic Versus Open Resection for Colorectal Liver Metastases: The OSLO-COMET Randomized Controlled Trial.* Annals of Surgery, vol. 267, no. 2, page 199–207, February 2018. (Cited on page 3.)

[Friebe 2026] Michael H. Friebe. *Haptic and Palpation Sensing for Robotic Surgery: Engineering Perspectives on Design and Integration.* Sensors, vol. 26, no. 4, page 1126, February 2026. (Cited on page 2.)

[Garrido-Jurado *et al.* 2014] S. Garrido-Jurado, R. Muñoz-Salinas, F.J. Madrid-Cuevas and M.J. Marín-Jiménez. *Automatic generation and detection of highly reliable fiducial markers under occlusion.* Pattern Recognition, vol. 47, no. 6, page 2280–2292, June 2014. (Cited on page 19.)

[GE Healthcare 2023] GE Healthcare. *Making Waves in the OR: How a Specialized Ultrasound System Helps Surgeons Make Better Decisions — gehealthcare.com.* online, 2023. [Accessed 03-07-2026]. (Cited on page 6.)

[Golse *et al.* 2021] Nicolas Golse, Antoine Petit, Maïté Lewin, Eric Vibert and Stéphane Cotin. *Augmented reality during open liver surgery using a markerless non-rigid registration system*. J. Gastrointest. Surg., vol. 25, no. 3, pages 662–671, March 2021. (Cited on page 48.)

[Gummeson & Oskarsson 2023] Anna Gummeson and Magnus Oskarsson. *Robust and Accurate Cylinder Triangulation*. In Image Analysis, pages 451–466, Cham, 2023. Springer Nature Switzerland. (Cited on page 31.)

[Gummeson *et al.* 2022] Anna Gummeson, Johanna Engman, Kalle Åström and Magnus Oskarsson. *Fast and efficient minimal solvers for quadric based camera pose estimation*. In 2022 26th International Conference on Pattern Recognition (ICPR), pages 3973–3979, 2022. (Cited on page 31.)

[Hadramy *et al.* 2024] Sidaty El Hadramy, Juan Verde, Nicolas Padoy and Stéphane Cotin. *Towards Real-Time Vessel Guided Augmented Reality for Liver Surgery*. In 2024 IEEE International Symposium on Biomedical Imaging (ISBI), page 1–5. IEEE, May 2024. (Cited on page 49.)

[Hale *et al.* 2025] Joshua Hale, Kelsey R. Landrum, Chris Agala, Roberto A. Vidri, Elizabeth Gleeson and Michael T. LeCompte. *Minimally invasive vs. open radical cholecystectomy for gallbladder cancer: 30-day NSQIP outcomes analysis*. Surgical Endoscopy, vol. 39, no. 6, page 3873–3882, May 2025. (Cited on page 2.)

[Hallet *et al.* 2016] Julie Hallet, Kaitlyn Beyfuss, Riccardo Memeo, Paul J. Karanicolas, Jacques Marescaux and Patrick Pessaux. *Short and long-term outcomes of laparoscopic compared to open liver resection for colorectal liver metastases*. HepatoBiliary Surgery and Nutrition, vol. 5, no. 4, page 300–310, August 2016. (Cited on page 3.)

[Hanifati *et al.* 2026] Kirana Hanifati, Mohammad Alkhatib, Erol Ozgur, Emmanuel Buc, Bertrand Le Roy, Hestiasari Rante, Youcef Mezouar and Adrien Bartoli. *Hidden Tumour Visualization in Augmented Monocular Liver Laparoscopy*. Healthcare Technology Letters, vol. 13, no. 1, January 2026. (Cited on page 3.)

[Iglovikov & Shvets 2018] Vladimir I. Iglovikov and Alexey A. Shvets. *TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation*. ArXiv, vol. abs/1801.05746, 2018. (Cited on page 34.)

[Iglovikov *et al.* 2017] Vladimir I. Iglovikov, Sergey Mushinskiy and Vladimir Osin. *Satellite Imagery Feature Detection using Deep Convolutional Neural Network: A Kaggle Competition*. ArXiv, vol. abs/1706.06169, 2017. (Cited on page 34.)

[International Agency for Research on Cancer 2024] International Agency for Research on Cancer. Globocan 2022: Global cancer observatory. World Health Organization, Lyon, France, 2024. (Cited on page 3.)

[Jaffray 2005] B Jaffray. *Minimally invasive surgery.* Archives of Disease in Childhood, vol. 90, no. 5, pages 537–542, May 2005. (Cited on pages 2 and 27.)

[Jia *et al.* 2018] Chenyang Jia, Hongyu Li, Ningyuan Wen, Junhua Chen, Yonggang Wei and Bo Li. *Laparoscopic liver resection: a review of current indications and surgical techniques.* HepatoBiliary Surgery and Nutrition, vol. 7, no. 4, page 277–288, August 2018. (Cited on page 3.)

[Jung 2021] Haijo Jung. *Basic Physical Principles and Clinical Applications of Computed Tomography.* Progress in Medical Physics, vol. 32, no. 1, page 1–17, March 2021. (Cited on page 6.)

[Kaess *et al.* 2011] Michael Kaess, Hordur Johannsson, Richard Roberts, Viorela Ila, John J Leonard and Frank Dellaert. *iSAM2: Incremental smoothing and mapping using the Bayes tree.* The International Journal of Robotics Research, vol. 31, no. 2, page 216–235, December 2011. (Cited on page 57.)

[Kalantari *et al.* 2024a] Mohammad Mahdi Kalantari, Erol Ozgur, Mohammad Alkhatib, Emmanuel Buc, Bertrand Le Roy, Richard Modrzejewski, Youcef Mezouar and Adrien Bartoli. *LARLUS: laparoscopic augmented reality from laparoscopic ultrasound.* International Journal of Computer Assisted Radiology and Surgery, vol. 19, no. 7, page 1285–1290, April 2024. (Cited on pages 69 and 70.)

[Kalantari *et al.* 2024b] Mohammad Mahdi Kalantari, Erol Ozgur, Mohammad Alkhatib, Emmanuel Buc, Bertrand Le Roy, Richard Modrzejewski, Youcef Mezouar and Adrien Bartoli. *Markerless Ultrasound Probe Pose Estimation in Mini-Invasive Surgery.* IEEE International Conference on Robotics and Automation, ICRA, 2024. (Cited on page 70.)

[Kallasi *et al.* 2014] Fabjan Kallasi, Fabio Oleari, Marco Bottioni, Dario Lodi Rizzini and Stefano Caselli. *Object Detection and Pose Estimation Algorithms for Underwater Manipulation.* 2014. (Cited on page 31.)

[Kingma & Ba 2017] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*, 2017. (Cited on page 34.)

[Koo *et al.* 2017] Bongjin Koo, Erol Ozgur, Bertrand Le Roy, Emmanuel Buc and Adrien Bartoli. *Deformable Registration of a Preoperative 3D Liver Volume to a Laparoscopy Image Using Contour and Shading Cues.* In International Conference on Medical Image Computing and Computer-Assisted Intervention, 2017. (Cited on pages 70 and 79.)

[Kurmann *et al.* 2017] Thomas Kurmann, Pablo Marquez Neila, Xiaofei Du, Pascal Fua, Danail Stoyanov, Sebastian Wolf and Raphael Sznitman. *Simultaneous Recognition and Pose Estimation of Instruments in Minimally Invasive Surgery.* In Medical Image Computing and Computer-Assisted Intervention - MICCAI 2017, pages 505–513, Cham, 2017. Springer International Publishing. (Cited on pages 30 and 31.)

[Labrunie *et al.* 2022] M Labrunie, M Ribeiro, F Mourthadhoi, C Tilmant, B Le Roy, E Buc and A Bartoli. *Automatic preoperative 3d model registration in laparoscopic liver resection.* IJCARS, vol. 17, no. 8, pages 1429–1436, August 2022. (Cited on pages 70, 71, 73, 76, 78, 79 and 81.)

[Labrunie *et al.* 2024] Mathieu Labrunie, Daniel Pizarro, Christophe Tilmant and Adrien Bartoli. *Automatic 3D/2D Deformable Registration in Minimally Invasive Liver Resection using a Mesh Recovery Network.* In Medical Imaging with Deep Learning, volume 227 of *Proceedings of Machine Learning Research*, pages 1104–1123, PMLR, 10–12 Jul 2024. PMLR. (Cited on pages 69, 70, 73, 77, 78 and 79.)

[Labrunie 2024] Mathieu Labrunie. *Contributions to the Automation of Preoperative 3D Model to 2D Image Registration in Mini-Invasive Liver Surgery: Primitive Detection, Pose Estimation, Registration and Anatomical Modelling.* PhD thesis, Université Clermont Auvergne, 2024. (Cited on pages 6 and 81.)

[Lai *et al.* 2024] Tzu-Jung Lai, Campbell Roxburgh, Kathleen Anne Boyd and Janet Bouttell. *Clinical effectiveness of robotic versus laparoscopic and open surgery: an overview of systematic reviews.* BMJ Open, vol. 14, no. 9, page e076750, September 2024. (Cited on page 2.)

[Laina *et al.* 2017] Iro Laina, Nicola Rieke, Christian Rupprecht, Josué Page Vizcaíno, Abouzar Eslami, Federico Tombari and Nassir Navab. *Concurrent Segmentation and Localization for Tracking of Surgical Instruments.* In Medical Image Computing and Computer-Assisted Intervention - MICCAI 2017, pages 664–672, Cham, 2017. Springer International Publishing. (Cited on pages 30 and 31.)

[Lau *et al.* 2019] Lung W. Lau, Xinyang Liu, William Plishker, Karun Sharma, Raj Shekhar and Timothy D. Kane. *Laparoscopic Liver Resection with Augmented Reality: A Preclinical Experience.* Journal of Laparoendoscopic & Advanced Surgical Techniques, vol. 29, no. 1, page 88–93, January 2019. (Cited on page 9.)

[Lauterbur 1973] P. C. Lauterbur. *Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance.* Nature, vol. 242, no. 5394, page 190–191, March 1973. (Cited on page 7.)

[Le Roy *et al.* 2019] B. Le Roy, E. Ozgur, B. Koo, E. Buc and A. Bartoli. *Augmented reality guidance in laparoscopic hepatectomy with deformable semi-automatic computed tomography alignment*. Journal of Visceral Surgery, vol. 156, no. 3, page 261–262, June 2019. (Cited on page 4.)

[Liang *et al.* 2025] Zekai Liang, Zih-Yun Chiu, Florian Richter and Michael C. Yip. *Differentiable Rendering-based Pose Estimation for Surgical Robotic Instruments*. In 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), page 20898–20905. IEEE, October 2025. (Cited on page 29.)

[Lin *et al.* 2023] Zhefan Lin, Chen Lei and Liangjing Yang. *Modern Image-Guided Surgery: A Narrative Review of Medical Image Processing and Visualization*. Sensors, vol. 23, no. 24, page 9872, December 2023. (Cited on page 8.)

[Liu *et al.* 2021] Xinyang Liu, William Plishker and Raj Shekhar. *Hybrid electromagnetic-ArUco tracking of laparoscopic ultrasound transducer in laparoscopic video*. Journal of Medical Imaging, vol. 8, no. 01, February 2021. (Cited on pages 48 and 70.)

[Liu *et al.* 2022] Kaidi Liu, Zijian Zhao, Pan Shi, Feng Li and He Song. *Real-time surgical tool detection in computer-aided surgery based on enhanced feature-fusion convolutional neural network*. Journal of Computational Design and Engineering, vol. 9, no. 3, page 1123–1134, May 2022. (Cited on pages 30 and 31.)

[Lu *et al.* 2023] Jingpei Lu, Florian Richter and Michael C. Yip. *Markerless Camera-to-Robot Pose Estimation via Self-Supervised Sim-to-Real Transfer*. In 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), page 21296–21306. IEEE, June 2023. (Cited on page 29.)

[Lubner *et al.* 2021] Meghan G Lubner, Lori Mankowski Gettle, David H Kim, Timothy J Ziemlewicz, Nirvikar Dahiya and Perry Pickhardt. *Diagnostic and procedural intraoperative ultrasound: technique, tips and tricks for optimizing results*. The British Journal of Radiology, vol. 94, no. 1121, April 2021. (Cited on page 5.)

[Maier-Hein *et al.* 2017] Lena Maier-Hein, Swaroop S. Vedula, Stefanie Speidel, Nassir Navab, Ron Kikinis, Adrian Park, Matthias Eisenmann, Hubertus Feussner, Germain Forestier, Stamatia Giannarou, Makoto Hashizume, Darko Katic, Hannes Kenngott, Michael Kranzfelder, Anand Malpani, Keno März, Thomas Neumuth, Nicolas Padoy, Carla Pugh, Nicolai Schoch, Danail Stoyanov, Russell Taylor, Martin Wagner, Gregory D. Hager and Pierre Jannin. *Surgical data science for next-generation interventions*. Nature Biomedical Engineering, vol. 1, no. 9, page 691–696, September 2017. (Cited on page 5.)

[Makki *et al.* 2023] Karim Makki, Kilian Chandelon and Adrien Bartoli. *Elliptical specularity detection in endoscopy with application to normal reconstruction.* International Journal of Computer Assisted Radiology and Surgery, vol. 18, no. 7, page 1323–1328, May 2023. (Cited on page 5.)

[Mhiri *et al.* 2024] Islem Mhiri, Daniel Pizarro and Adrien Bartoli. *Neural patient-specific 3D-2D registration in laparoscopic liver resection.* IJCARS, July 2024. (Cited on pages 9, 69, 70 and 79.)

[Mohareri *et al.* 2013] Omid Mohareri, Caitlin Schneider, Troy K. Adebar, Mike C. Yip, Peter Black, Christopher Y. Nguan, Dale Bergman, Jonathan Seroger, Simon DiMaio and Septimiu E. Salcudean. *Ultrasound-Based Image Guidance for Robot-Assisted Laparoscopic Radical Prostatectomy: Initial in-vivo Results.* In Information Processing in Computer-Assisted Interventions, pages 40–50, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg. (Cited on page 28.)

[Montaña-Brown *et al.* 2021] Nina Montaña-Brown, João Ramalhinho, Moustafa Allam, Brian Davidson, Yipeng Hu and Matthew J. Clarkson. *Vessel segmentation for automatic registration of untracked laparoscopic ultrasound to CT of the liver.* International Journal of Computer Assisted Radiology and Surgery, may 2021. (Cited on pages 30, 67 and 70.)

[Montaña-Brown *et al.* 2022] Nina Montaña-Brown, João Ramalhinho, Bongjin Koo, Moustafa Allam, Brian Davidson, Kurinchi Gurusamy, Yipeng Hu and Matthew J. Clarkson. *Towards Multi-modal Self-supervised Video and Ultrasound Pose Estimation for Laparoscopic Liver Surgery.* In Simplifying Medical Ultrasound, pages 183–192, Cham, 2022. Springer International Publishing. (Cited on pages 30, 49 and 70.)

[Morise & Wakabayashi 2017] Zenichi Morise and Go Wakabayashi. *First quarter century of laparoscopic liver resection.* World Journal of Gastroenterology, vol. 23, no. 20, page 3581, 2017. (Cited on page 3.)

[Morise *et al.* 2015] Zenichi Morise, Ruben Ciria, Daniel Cherqui, Kuo-Hsin Chen, Giulio Belli and Go Wakabayashi. *Can we expand the indications for laparoscopic liver resection? A systematic review and meta-analysis of laparoscopic liver resection for patients with hepatocellular carcinoma and chronic liver disease.* Journal of Hepato-Biliary-Pancreatic Sciences, vol. 22, no. 5, page 342–352, February 2015. (Cited on page 3.)

[Murray 1994] Richard M Murray. A mathematical introduction to robotic manipulation. CRC Press, Boca Raton, FL, March 1994. (Cited on page 53.)

[Nainamalai *et al.* 2025] Varatharajan Nainamalai, Håvard Bjørke Jenssen, Luca Boretto, Nikhil Pramod Kumar, Andreas Westenvik Espinoza, Egidijus Pelanis, Bård I. Røsok, Ole Jakob Elle, Ingrid Schrøder Hansen, Bjørn Edwin and Åsmund Avdem Fretland. *Morphological changes on the human liver during minimally invasive surgery: Implications for image-guided interventions and surgical navigation*. Surgical Endoscopy, vol. 40, no. 2, page 1458–1468, November 2025. (Cited on page 4.)

[Noble & Boukerroui 2006] J.A. Noble and D. Boukerroui. *Ultrasound image segmentation: a survey*. IEEE Transactions on Medical Imaging, vol. 25, no. 8, page 987–1010, August 2006. (Cited on page 5.)

[Ntourakis *et al.* 2015] Dimitrios Ntourakis, Ricardo Memeo, Luc Soler, Jacques Marescaux, Didier Mutter and Patrick Pessaux. *Augmented Reality Guidance for the Resection of Missing Colorectal Liver Metastases: An Initial Experience*. World Journal of Surgery, vol. 40, no. 2, page 419–426, August 2015. (Cited on page 48.)

[NVIDIA 2025a] NVIDIA. *NVIDIA TensorRT Developer Guide*. `https://docs.nvidia.com/deeplearning/tensorrt/developer-guide/index.html`, 2025. Accessed: 2026-04-25. (Cited on page 35.)

[NVIDIA 2025b] NVIDIA. *NVIDIA TensorRT Documentation*. `https://docs.nvidia.com/deeplearning/tensorrt/index.html`, 2025. Accessed: 2026-04-25. (Cited on page 35.)

[Olthof *et al.* 2026] Karin A Olthof, Lisanne P J Venix, Maaike Pruijt, Anne G den Hartog, Niels F M Kok, Theo J M Ruers, Matteo Fusaglia and Koert F D Kuhlmann. *Toward augmented reality in laparoscopic liver surgery using electromagnetic tracking: a clinical feasibility study*. Surg. Endosc., vol. 40, no. 4, pages 3538–3547, April 2026. (Cited on page 7.)

[Owen *et al.* 2016] Joseph W. Owen, Kathryn J. Fowler, Maria B. Doyle, Nael E. Saad, David C. Linehan and William C. Chapman. *Colorectal liver metastases: disappearing lesions in the era of Eovist hepatobiliary magnetic resonance imaging*. HPB, vol. 18, no. 3, page 296–303, March 2016. (Cited on page 48.)

[Ozgur *et al.* 2026] Erol Ozgur, Mohammad Alkhatib, Youcef Mezouar and Adrien Bartoli. *Reconstructing a Sphere and the Camera Focal Length from a Single View by Fitting Planes*. International Journal of Computer Vision, vol. 134, no. 4, March 2026. (Cited on page 32.)

[Park *et al.* 2024] Jihun Park, Jiuk Hong, Jihun Yoon, Bokyung Park, Min-Kook Choi and Heechul Jung. Towards precise pose estimation in robotic surgery: Introducing occlusion-aware loss, page 639–648. Springer Nature Switzerland, 2024. (Cited on page 29.)

[Peek *et al.* 2024] J. J. Peek, X. Zhang, K. Hildebrandt, S. A. Max, A. H. Sadeghi, A. J. J. C. Bogers and E. A. F. Mahtab. *A novel 3D image registration* *technique for augmented reality vision in minimally invasive thoracoscopic pulmonary segmentectomy.* International Journal of Computer Assisted Radiology and Surgery, vol. 20, no. 4, page 787–795, December 2024. (Cited on page 9.)

[Pelanis *et al.* 2021] Egidijus Pelanis, Andrea Teatini, Benjamin Eigl, Alois Regensburger, Amilcar Alzaga, Rahul Prasanna Kumar, Tobias Rudolph, Davit L. Aghayan, Carina Riediger, Niclas Kvarnström, Ole Jakob Elle and Bjørn Edwin. *Evaluation of a novel navigation platform for laparoscopic liver surgery with organ deformation compensation using injected fiducials.* Medical Image Analysis, vol. 69, page 101946, 2021. (Cited on page 48.)

[Rabbani *et al.* 2022a] N. Rabbani, L. Calvet, Y. Espinel, B. Le Roy, M. Ribeiro, E. Buc and A. Bartoli. *A methodology and clinical dataset with ground-truth to evaluate registration accuracy quantitatively in computer-assisted Laparoscopic Liver Resection.* Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, vol. 10, no. 4, pages 441–450, 2022. (Cited on pages 9, 14, 15, 16, 24, 25, 44, 48, 70, 78, 79 and 80.)

[Rabbani *et al.* 2022b] Navid Rabbani, Callyane Seve, Nicolas Bourdel and Adrien Bartoli. *Video-based Computer-aided Laparoscopic Bleeding Management: a Space-time Memory Neural Network with Positional Encoding and Adversarial Domain Adaptation.* In Proceedings of The 5th International Conference on Medical Imaging with Deep Learning, volume 172 of *Proceedings of Machine Learning Research*, pages 961–974. PMLR, 06–08 Jul 2022. (Cited on page 5.)

[Ramalhinho *et al.* 2019] João Ramalhinho, Henry Tregidgo, Moustafa Allam, Nikolina Travlou, Kurinchi Gurusamy, Brian Davidson, David Hawkes, Dean Barratt and Matthew J. Clarkson. *Registration of Untracked 2D Laparoscopic Ultrasound Liver Images to CT Using Content-Based Retrieval and Kinematic Priors.* In Smart Ultrasound Imaging and Perinatal, Preterm and Paediatric Image Analysis, pages 11–19, Cham, 2019. Springer International Publishing. (Cited on pages 30 and 70.)

[Ramalhinho *et al.* 2021] João Ramalhinho, Henry F. J. Tregidgo, Kurinchi Gurusamy, David J. Hawkes, Brian Davidson and Matthew J. Clarkson. *Registration of Untracked 2D Laparoscopic Ultrasound to CT Images of the Liver Using Multi-Labelled Content-Based Image Retrieval.* IEEE Transactions on Medical Imaging, vol. 40, no. 3, pages 1042–1054, 2021. (Cited on page 30.)

[Ramalhinho *et al.* 2022] João Ramalhinho, Bongjin Koo, Nina Montaña-Brown, Shaheer U. Saeed, Ester Bonmati, Kurinchi Gurusamy, Stephen P. Pereira, Brian Davidson, Yipeng Hu and Matthew J. Clarkson. *Deep hashing for global registration of untracked 2D laparoscopic ultrasound to CT.* International Journal of Computer Assisted Radiology and Surgery, vol. 17, no. 8, page 1461–1468, April 2022. (Cited on page 70.)

[Ramalhinho *et al.* 2023] João Ramalhinho, Soojeong Yoo, Thomas Dowrick, Bongjin Koo, Murali Somasundaram, Kurinchi Gurusamy, David J. Hawkes, Brian Davidson, Ann Blandford and Matthew J. Clarkson. *The value of Augmented Reality in surgery — A usability study on laparoscopic liver surgery.* Medical Image Analysis, vol. 90, page 102943, December 2023. (Cited on page 3.)

[Reich *et al.* 1991] H Reich, F Mcglynn, J Decaprio and R Budin. *Laparoscopic excision of benign liver lesions.* Obstetrics and gynecology, vol. 78, no. 5, pages 956–958, 1991. (Cited on page 3.)

[Ribeiro *et al.* 2024] Mathieu Ribeiro, Yamid Espinel, Navid Rabbani, Bruno Pereira, Adrien Bartoli and Emmanuel Buc. *Augmented reality guided laparoscopic liver resection: A phantom study with intraparenchymal tumors.* The Journal of surgical research, vol. 296, pages 612–620, April 2024. (Cited on pages 14, 16, 17, 22 and 77.)

[Ricciardi *et al.* 2024] Rocco Ricciardi, Usha Seshadri-Kreaden, Ana Yankovsky, Douglas Dahl, Hugh Auchincloss, Neera M. Patel, April E. Hebert and Valena Wright. *The COMPARE Study: Comparing Perioperative Outcomes of Oncologic Minimally Invasive Laparoscopic, da Vinci Robotic, and Open Procedures: A Systematic Review and Meta-analysis of the Evidence.* Annals of Surgery, vol. 281, no. 5, page 748–763, October 2024. (Cited on page 2.)

[Robu *et al.* 2018] Maria R Robu, João Ramalhinho, Stephen Thompson, Kurinchi Gurusamy, Brian Davidson, David Hawkes, Danail Stoyanov and Matthew J Clarkson. *Global rigid registration of CT to video in laparoscopic liver surgery.* International Journal of Computer Assisted Radiology and Surgery, vol. 13, no. 6, pages 947–956, June 2018. (Cited on page 70.)

[Ronneberger *et al.* 2015] Olaf Ronneberger, Philipp Fischer and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation.* In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, pages 234–241, Cham, 2015. Springer International Publishing. (Cited on page 34.)

[Rumgay *et al.* 2022] Harriet Rumgay, Melina Arnold, Jacques Ferlay, Olufunmilayo Lesi, Citadel J. Cabasag, Jérôme Vignat, Mathieu Laversanne, Katherine A. McGlynn and Isabelle Soerjomataram. *Global burden of primary liver cancer in 2020 and predictions to 2040.* Journal of Hepatology, vol. 77, no. 6, page 1598–1606, December 2022. (Cited on page 3.)

[Santambrogio *et al.* 2022] Roberto Santambrogio, Maurizio Vertemati, Edgardo Picardi and Marco Antonio Zappa. *Planning the treatment: preoperative 3D reconstruction.* Laparoscopic Surgery, vol. 6, no. 0, 2022. (Cited on page 5.)

[Sarikaya *et al.* 2017] Duygu Sarikaya, Jason J. Corso and Khurshid A. Guru. *Detection and Localization of Robotic Tools in Robot-Assisted Surgery Videos Using Deep Neural Networks for Region Proposal and Detection.* IEEE Transactions on Medical Imaging, vol. 36, no. 7, pages 1542–1549, 2017. (Cited on pages 30 and 31.)

[Schneider *et al.* 2021] C. Schneider, M. Allam, D. Stoyanov, D.J. Hawkes, K. Gurusamy and B.R. Davidson. *Performance of image guided navigation in laparoscopic liver surgery – A systematic review.* Surgical Oncology, vol. 38, page 101637, September 2021. (Cited on page 9.)

[Shvets *et al.* 2018] Alexey A. Shvets, Alexander Rakhlin, Alexandr A. Kalinin and Vladimir I. Iglovikov. *Automatic Instrument Segmentation in Robot-Assisted Surgery using Deep Learning.* In 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), pages 624–628, 2018. (Cited on page 34.)

[Singla *et al.* 2017] Rohit Singla, Philip Edgcumbe, Philip Pratt, Christopher Nguan and Robert Rohling. *Intra-operative ultrasound-based augmented reality guidance for laparoscopic surgery.* Healthcare Technology Letters, vol. 4, no. 5, pages 204–209, September 2017. (Cited on page 28.)

[Song *et al.* 2015] Yi Song, Johannes Totz, Steve Thompson, Stian Johnsen, Dean Barratt, Crispin Schneider, Kurinchi Gurusamy, Brian Davidson, Sébastien Ourselin, David Hawkes and Matthew J. Clarkson. *Locally rigid, vessel-based registration for laparoscopic liver surgery.* International Journal of Computer Assisted Radiology and Surgery, vol. 10, no. 12, page 1951–1961, June 2015. (Cited on page 70.)

[Strasberg 2005] Steven M. Strasberg. *Nomenclature of hepatic anatomy and resections: a review of the Brisbane 2000 system.* Journal of Hepato-Biliary-Pancreatic Surgery, vol. 12, no. 5, page 351–355, October 2005. (Cited on page 3.)

[Suzuki & Abe 1985] Satoshi Suzuki and Keiichi Abe. *Topological structural analysis of digitized binary images by border following.* Comput. Vis. Graph. Image Process., vol. 30, pages 32–46, 1985. (Cited on page 35.)

[The Liver Center] The Liver Center. *Types of Surgery / Open surgery or Laparoscopy, Gastroscopy.* `https://www.livercentrewa.com.au/treatments/types-of-surgery-perth/`. [Accessed 29-06-2026]. (Cited on page 2.)

[Toth & Hajder 2023] T. Toth and L. Hajder. *A Minimal Solution for Image-Based Sphere Estimation.* International Journal of Computer Vision, vol. 131, pages 1428–1447, 2023. (Cited on pages 32, 41 and 42.)

[Traynor *et al.* 1988] O Traynor, D Castaing and H Bismuth. *Peroperative ultrasonography in the surgery of hepatic tumours.* British Journal of Surgery, vol. 75, no. 3, pages 197–202, March 1988. (Cited on page 47.)

[Vazquez *et al.* 2025] Valerie K. Vazquez, Laura Mena Albors, Huazhi Liu and Darwin Ang. *Outcomes Comparing 19, 769 Patients Who Underwent Robotic, Laparoscopic, or Open Colorectal Surgery.* JSLS: Journal of the Society of Laparoendoscopic Surgeons,, vol. 29, no. 3, page e2025.00038, 2025. (Cited on page 2.)

[Viganò *et al.* 2016] L Viganò, G Torzilli, M Cimino, K Imai, E Vibert, M Donadon, D Castaing and R Adam. *Drop-out between the two liver resections of two-stage hepatectomy. Patient selection or loss of chance?* European Journal of Surgical Oncology, vol. 42, no. 9, pages 1385–1393, September 2016. (Cited on page 48.)

[Wakabayashi *et al.* 2015] Go Wakabayashi, Daniel Cherqui, David A Geller, Joseph F Buell, Hironori Kaneko, Ho Seong Han, Horacio Asbun, Nicholas ORourke, Minoru Tanabe, Alan J Koffron, Allan Tsung, Olivier Soubrane, Marcel Autran Machado, Brice Gayet, Roberto I Troisi, Patrick Pessaux, Ronald M Van Dam, Olivier Scatton, Mohammad Abu Hilal, Giulio Belli, Choon Hyuck David Kwon, Bjørn Edwin, Gi Hong Choi, Luca Antonio Aldrighetti, Xiujun Cai, Sean Cleary, Kuo-Hsin Chen, Michael R Schön, Atsushi Sugioka, Chung-Ngai Tang, Paulo Herman, Juan Pekolj, Xiao-Ping Chen, Ibrahim Dagher, William Jarnagin, Masakazu Yamamoto, Russell Strong, Palepu Jagannath, Chung-Mau Lo, Pierre-Alain Clavien, Norihiro Kokudo, Jeffrey Barkun and Steven M Strasberg. *Recommendations for laparoscopic liver resection: a report from the second international consensus conference held in Morioka.* Annals of Surgery, vol. 261, no. 4, pages 619–629, April 2015. (Cited on page 3.)

[Wang *et al.* 2015] Defeng Wang, Diya Ma, Matthew Lun Wong and Yì Xiáng J. Wang. *Recent advances in surgical planning & navigation for tumor biopsy and resection.* Quantitative Imaging in Medicine and Surgery, vol. 5, no. 5, 2015. (Cited on page 8.)

[Wang *et al.* 2022] Risheng Wang, Tao Lei, Ruixia Cui, Bingtao Zhang, Hongying Meng and Asoke K. Nandi. *Medical image segmentation using deep learning: A survey.* IET Image Processing, vol. 16, no. 5, page 1243–1267, January 2022. (Cited on page 5.)

[Wong *et al.* 2008] Kwan-Yee K. Wong, Dirk Schnieders and Shuda Li. *Recovering Light Directions and Camera Poses from a Single Sphere.* In Computer Vision – ECCV 2008, pages 631–642, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg. (Cited on page 31.)

[Xiao *et al.* 2024] Zhen-Kun Xiao, Yong-Hong Duan, Xin-Yu Mao, Ri-Chu Liang, Min Zhou and Yong-Mei Yang. *Traditional craniotomy versus current minimally invasive surgery for spontaneous supratentorial intracerebral haemorrhage: A propensity-matched analysis.* World Journal of Radiology, vol. 16, no. 8, page 317–328, August 2024. (Cited on page 2.)

[Xu *et al.* 2025] Haozheng Xu, Alistair Weld, Chi Xu, Alfie Roddan, João Cartucho, Mert Asim Karaoglu, Alexander Ladikos, Yangke Li, Yiping Li, Daiyun Shen, Geonhee Lee, Seyeon Park, Jongho Shin, Lucy Fothergill, Dominic Jones, Pietro Valdastri, Duygu Sarikaya and Stamatia Giannarou. *SurgRIPE challenge: Benchmark of surgical robot instrument pose estimation.* Medical Image Analysis, vol. 105, page 103674, October 2025. (Cited on page 29.)

[Ye *et al.* 2016] Menglong Ye, Lin Zhang, Stamatia Giannarou and Guang-Zhong Yang. *Real-Time 3D Tracking of Articulated Tools for Robotic Surgery.* In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016, pages 386–394, Cham, 2016. Springer International Publishing. (Cited on pages 29 and 31.)

[Zenteno *et al.* 2019] Omar Zenteno, Sylvie Treuillet and Yves Lucas. *3D Cylinder Pose Estimation by Maximization of Binary Masks Similarity: A simulation Study for Multispectral Endoscopy Image Registration.* In VISIGRAPP, 2019. (Cited on page 30.)

[Zenteno *et al.* 2021] Omar Zenteno, Sylvie Treuillet and Yves Lucas. *Pose estimation of a markerless fiber bundle for endoscopic optical biopsy.* Journal of Medical Imaging, vol. 8, no. 02, mar 2021. (Cited on pages 30 and 31.)

[Zhang & Zhang 2020] Zhanpeng Zhang and Kaipeng Zhang. *FarSee-Net: Real-Time Semantic Segmentation by Efficient Multi-scale Context Aggregation and Feature Space Super-resolution*, 2020. (Cited on page 34.)