Machine Stereo Vision for Medical Image Registration

Andrew David Speers

A Dissertation submitted to
the Faculty of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

Graduate Program in
Electrical Engineering and Computer Science
York University
Toronto, Ontario

September 2020

Abstract

Image guided liver surgery aims to enhance the precision of resection and ablation by providing fast localization of tumours and adjacent complex vasculature to improve oncologic outcome. This dissertation presents a novel end-to-end system for fast and accurate 3D surface reconstruction and motion estimation of the liver for alignment of intraoperative imagery with a preoperative volumetric scan. The system is designed and evaluated for application to liver surgery in an open setting, where open surgery is the dominant setting. The system is comprised of three key components: initialization, 3D surface recovery, and 3D motion estimation. Initialization is performed semi-automatically using a *Branch-and-Bound (BnB)* strategy to generate a set of globally optimal shape-based registration candidates from which the user can select a suitable initialization. 3D surface recovery is performed using a computationally efficient adaptive *Coarse-to-Fine (CTF)* stereo algorithm providing data-driven dense reconstructions in a computationally-efficient manner. A robust, 3D motion estimation technique based on interframe feature matching is then used to register a time series of reconstructions back to the initial frame of the sequence. The system has been evaluated empirically with reference to novel laboratory and intraoperative datasets, with results showing that performance is within tolerances expected for integration into *Surgical Navigation (SN)* systems.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Professor Richard Wildes. You encouraged me to take on a challenge and had the patience to support me through to the end. Your guidance and support has been integral for me getting to this point where I now find myself writing the acknowledgements section on my successfully defended dissertation. It has been a true pleasure working under your guidance.

I was fortunate to have several collaborators who enabled me to be able to apply this work in a clinical setting and who helped ensure that this work was targeting something that would be clinically relevant. I would like to thank my academic collaborators Professor Amber Simpson for instigating our collaboration with Memorial Sloan Kettering Cancer Center and for showing us the ropes of performing research in a clinical setting and Professor Burton Ma for our insightful discussions and helping to ensure that the best practices of the *CAI* community were represented in the evaluation of this work. I would like to thank our clinical collaborator Dr. William Jarnagin for allowing us to enter your operating room to collect data. Seeing you and your colleagues do the life saving work you do in such an exceptional manner was truly inspiring. I would also like to thank Dr. Sharifa Himidan for hosting us at The Hospital for Sick Children when we were figuring out the logistical issues of using the system in a surgical setting and for helping us to expedite obtaining *CT* scans for our laboratory data when conference deadlines were looming. To all the surgeons, principle investigators, surgical residents and graduate students who took part in our surgical cases, helping

to collect intraoperative data and in the expert segmentation of preoperative *CT* scans, thank you for enabling this work.

I would like to thank my supervisory committee members, Professor John Tsotsos and Professor Kosta Derpanis, for their thoughtful feedback, support, and encouragement throughout my doctoral studies. I would also like to express my gratitude to the members who joined to serve on my examining committee: internal member, Professor Gunho Sohn, external examiner, Professor Michael Greenspan, and chair, Professor Michael Brown. I greatly appreciated the feedback on my dissertation made by all members of the committee, the results of which are found throughout this document. I am also thankful for the supportive, yet challenging, environment everyone provided for my defence.

I would like to thank my lab mates in the Computer Vision Lab and my peers in the Electrical Engineering and Computer Science Department. The stimulating conversations about research, countless moments of mutual support, and the always needed light-hearted moments provided a great deal of value to my life as a graduate student. The friendships gained throughout pursuing my doctorate are something I am truly thankful for.

I have had the good fortune to have been born into a supportive family and to have grown an amazing network of friends and family along the way. My deepest gratitude goes out to my friends and family, especially Amandeep Maroke, Steven Castellucci and Mudita Kundra. To everyone who provided support or encouragement throughout this journey, thank you!

Finally, I would like to dedicate this work to my parents, Leslie and Herbert Speers. Of the many blessings in my life I am most fortunate to have been blessed with parents who provided endless love, support and encouragement. With love, thank you both! This one is for us!

Contents

List of Tables

List of Figures

Acronyms

1PR 1-Point Random Sample Consensus.

AR Augmented Reality.

BA Bundle Adjustment.

BnB Branch-and-Bound.

BRW Brown-Roberts-Wells.

CAD Computer-Aided Design.

CAI Computer-Aided Intervention.

CAM Computer-Aided Manufacturing.

CAS Computer-Aided Surgery.

CBCT Cone Beam Computed Tomography.

CCD Charge-Coupled Device.

CNN Convolutional Neural Network.

CPD Coherent Point Drift.

CPU Central Processing Unit.

CRW Cosman-Roberts-Wells.

CT Computed Tomography.

CTF Coarse-to-Fine.

DIR Deformable Image Registration.

DLT Direct Linear Transformation.

DNN Deep Neural Network.

DOF Degrees of Freedom.

DP Dynamic Programming.

EGI Extended Gaussian Image.

EKF Extended Kalman Filter.

ENT Ear, Nose, and Throat.

FAST Features from Accelerated Segment Test.

FCN Fully Convolutional Network.

FDA Food and Drug Administration.

FoV Field-of-View.

FPFH Fast Point Feature Histogram.

FPS Frames Per Second.

GN Guided Navigation.

GO-ICP Globally Optimal Iterative Closest Point.

GPU Graphics Processing Unit.

GUI Graphical User Interface.

HMA Hierarchical MultiAffine.

HRM Hybrid Recursive Matching.

ICP Iterative Closest Point.

IGS Image Guided Surgery.

IL Intraoperative Localization.

IoU Intersection over Union.

LM Levenberg-Marquardt.

LPN Laparoscopic Partial Nephrectomy.

LRS Laser Range Scanning.

MC-SLAM Motion Compensated, Simultaneous Localization and Mapping.

MIS Minimally Invasive Surgery.

ML Machine Learning.

MLS Moving Least Squares.

MRI Magnetic Resonance Imaging.

MSE Mean Squared Error.

MSKCC Memorial Sloan Kettering Cancer Center.

NCC Normalized Cross Correlation.

NRSfM Non-Rigid Structure-from-Motion.

OR Operating Room.

ORB Oriented Features from Accelerated Segment Test and Rotated Binary Robust Independent Elementary Features.

P3P Perspective-Three-Point.

PCA Principal Component Analysis.

R-CNN Region-based Convolutional Neural Network.

RANSAC Random Sample Consensus.

RMS Root Mean Square.

RMSE Root Mean Square Error.

ROI Region Of Interest.

RPN Region Proposal Network.

SAD Sum of Absolute Differences.

SfM Structure-from-Motion.

SfS Shape-from-Shading.

SfT Shape-from-Template.

SGBM Semi-Global Block Matching.

SIFT Scale-Invariant Feature Transform.

SLAM Simultaneous Localization and Mapping.

SN Surgical Navigation.

SOR Successive Over-Relaxation.

SRE Surface Registration Error.

SURF Speeded Up Robust Features.

SVD Singular Value Decomposition.

SW Sliding-Window.

TRE Target Registration Error.

TSDF Truncated Signed Distance Function.

US Ultrasound.

VGG Visual Geometry Group.

ZNCC Zero-Mean Normalized Cross-Correlation.

Chapter 1

Introduction

1.1 Motivation

Medical imaging now plays an important role in the operating room in a field known as *Image Guided Surgery (IGS)* [138]. *IGS* techniques aim to aid surgical teams in operating within surgical tolerances by providing them with additional information beyond what is otherwise available about the patient's anatomy during the operation. *IGS* techniques can be as straightforward as directly deploying an imaging system, such as intraoperative *Ultrasound (US)*, that provides the surgical team the ability to view projections of subsurface structures that can guide actions during a surgery. Alternatively, more advanced *IGS* techniques use intraoperative imaging modalities, ranging from *US* and visible monocular or stereo vision systems all the way to volumetric scanners such as *Cone Beam Computed Tomography (CBCT)* or open *Magnetic Resonance Imaging (MRI)*, and register them to the preoperative imagery to provide surgical teams with a means of readily relating the information acquired preoperatively with the patient's anatomy in its intraoperative

state.

The more advanced forms of image guidance are often also referred to using terms such as *Computer-Aided Surgeries (CASs)*, *Computer-Aided Interventions (CAIs)*, *Guided Navigation (GN)*, *Surgical Navigation (SN)*, and *Intraoperative Localization (IL)*, primarily depending on the branch of the literature where the work appears. Nevertheless, all of these terms fall under the umbrella of *IGS*. For the purposes of this dissertation, the term *Surgical Navigation* (*SN*) will be used to denote systems that attempt to merge or register preoperative imagery to the intraoperative state of the patient to provide additional vital information to the surgical team.

Neurosurgery was the first surgical specialty to make use of *SN* techniques as a part of patient care. Early *SN* applications to neurosurgery were made possible by assuming that the spatial relationship between the brain tissue and surrounding bony (i.e., rigid) landmarks remained constant. With this assumption in place, initial *SN* systems made use of a stereotactic frame that was rigidly affixed to the patient's skull prior to obtaining preoperative imagery and would remain in place during surgery [190]. The transformation between the stereotactic frame and the preoperative imagery would be obtained through the localization of fiducial markers on the stereotactic frame in the preoperative imagery. In the case of the commonly used *Brown-Roberts-Wells (BRW)* and *Cosman-Roberts-Wells (CRW)* stereotactic frames, these fiducials came in the form of three oblique rods that extended out of the stereotactic frame in what was called a "localization frame". The relationship between the frame and the lesion/tumour observed in the preoperative imagery would be assumed to be rigid allowing the identification of landmarks on the stereotactic frame to guide the instruments to targets in the brain. Newer image-

guided techniques, also known as frameless stereotaxy, make use of less invasive fiducial markers for registration and external trackers for navigation of hand-held instruments [147, 57, 58]. These techniques notably still rely on the predominantly rigid relationship between the fiducial, skull, and brain tissue. With knowledge of the transformation between the stereotactic frame and the preoperative imagery, computers could be used to plan straight-line paths into areas of the brain using the stereotactic coordinate system.

Integration of rigid *SN* techniques have made their way into other surgical fields. Otolaryngology (also known as *Ear, Nose, and Throat (ENT)*) [25, 204, 122, 43, 102], orthopedics [201], maxillofacial surgery [218], and dental surgery [45] have all benefited from advances in *SN* technology. Similar to neurology, these applications of rigid *SN* techniques are possible due to the (semi-) rigidity of the surgical site, which is predominantly due to the close proximity of bones that limit the amount of deformation of the tissues of interest [14].

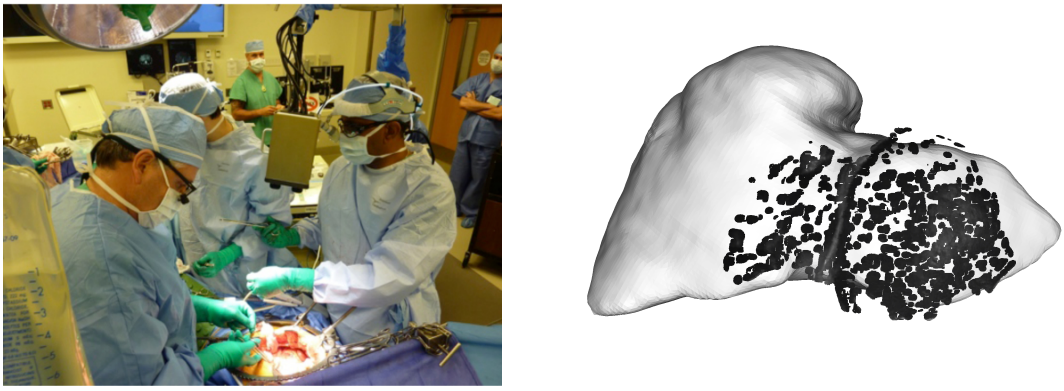
Surgical areas still awaiting mass adoption of *SN* techniques are those where the tissues of interest do not have a relatively constant spatial relationship with rigid structures, i.e., abdominal surgery. Here, non-rigid deformations may occur not only between preoperative and intraoperative imagery, but also between intraoperative scans acquired at different times, particularly when considering the spatial relationship relating the organ of interest to nearby bony structures.

Nevertheless, interest in the medical community in *SN* techniques for soft tissue surgery has grown steadily in recent years. One such area of interest is the use of *SN* techniques during liver resection tasks. Liver resection is the only potentially curative therapy for liver cancer, but often represents a surgical challenge due to the location of tumours throughout complex, delicate vasculature [155, 163]. To

date, *SN* applications to the liver have shown utility in specific surgical settings. *SN* applications to the liver have proved helpful in identifying small sonographically occult colorectal liver metastases (i.e., tumours that were found in preoperative *Computed Tomography (CT)*, but become difficult to see using intraoperative *US*), particularly after treatment with chemotherapy (often used to downstage the disease and allow for resection) [98, 86, 133]. *SN* systems have been used to predict accurate resection planes and remnant volumes [174]. When analyzing overall surgical outcomes, *SN* applications to liver surgery have had limited reports suggesting current techniques are able to aid in difficult ablations [87], able to provide slight reductions in incomplete resections [24], and have provided a benefit to individual patients undergoing complicated liver procedures [134]. Initial results suggest that by imaging the liver in real-time and providing the surgeon continuous feedback regarding the localization of tumours relative to adjacent major vasculature, the likelihood of an incomplete resection or inadvertent liver injury has the potential to be reduced [137, 134, 175]. However, similar to the state of neurological image guidance systems, no randomized controlled clinical trials have been performed to accurately assess the impact of current image guidance systems on patient outcomes in liver surgery [175]. The absence of such a study is primarily due to the fact that determining relevant endpoints for such a study that provides a sufficient number of patients is difficult [175].

1.2 Problem structure

The goal of a *Surgical Navigation (SN)* system is to achieve and maintain the registration of an intraoperative video source to the volumetric scan obtained



(a) (b)

Figure 1.1: Surgical Navigation (SN) Application to Liver Surgery. (a) Shows a view of the surgical theatre during an open liver resection surgery at *Memorial Sloan Kettering Cancer Center (MSKCC)*. The gold box positioned over the patient contains a stereo camera. Synchronized stereo video is used to reconstruct a time series of 3D surface models of the patient's liver, capturing a selected portion of the intraoperative state of the patient's anatomy. (b) Shows the subsequent registration of the surface model to the preoperative volumetric image of the patient's liver. The light points in the background of the image are the result of segmenting the outer surface of the patient's liver from a preoperative *CT* scan. The dark points in the foreground are the result of a stereo reconstruction performed on a single frame during the same patient's operation. Details of the algorithm leading from the input to the output are provided in Figure 3.1 and elaborated in Chapter 3.

preoperatively of the same anatomical region allowing for the two sources of data to be readily co-referenced by the surgical team. The majority of solutions to this task are shape-based approaches. In this case, to provide a common reference frame between the preoperative and intraoperative imagery, two sub-problems, each of which push state-of-the-art in computer vision, need to be solved. First, shape-based approaches start by estimating a 3D surface model of a selected portion of the patient's anatomy from the intraoperative imagery. Second, the intraoperative surface and preoperative volumetric model are registered. This registration is maintained on incoming imagery to keep the relationship of the two data sources up to date. A pictorial overview is provided in Figure 1.1.

There are three major challenges to realizing *Surgical Navigation (SN)* systems. First, a time series of 3D surface models must be recovered from an intraoperative video sequence even in the presence of sparse visual features and low texture surfaces that can be typical of internal organs, e.g., the liver. Moreover, estimation must be efficient enough to support real-time decision making in the operating theatre. Surface reconstruction approaches that can tolerate such conditions should be explored. Second, initial alignment between the recovered surface model and preoperative scan must be accomplished taking into account the cross modality differences between the intraoperative and preoperative imagery, i.e., different imaging modalities yield different visual appearances. Currently, initial alignment typically is performed by an expert human operator (e.g., a surgeon); however, this approach leads to interruption of the surgical work flow and a resulting accuracy that is highly user-dependent. Alternative approaches should be explored that lessen or remove the need for human intervention. Third, to support ongoing *Surgical Navigation (SN)* during surgery, the initial registration must be updated and maintained as incoming intraoperative imagery is acquired. Extant solutions to this challenge typically rely on repeated surface-to-surface registration of the recovered surface estimates; however, this approach often is susceptible to drift and thereby loss of physically meaningful registration beyond initialization. Alternative approaches that can maintain physically meaningful registration across a time-series of intraoperative acquisitions should be explored. Additionally, while not a challenge inherent to *SN* systems, advances in development of such systems currently are hampered by lack of controlled laboratory datasets for systematic evaluation.

1.3 Contributions

The primary contribution of this work is a novel end-to-end system for fast surface reconstruction and motion estimation of the liver for alignment of intraoperative imagery with a preoperative *CT* scan. Specifically, we deem this system to be "end-to-end" as the system implements a fully functional *Surgical Navigation (SN)* pipeline that relies on manual input for initialization purposes only and requires no subsequent user intervention. The current instantiation of this system is designed for use during intraoperative planning/exploratory phases in an open surgical setting. The targeted use case allows for the system to aid the surgical team with relocalizing subsurface structures of interest intraoperatively while accounting for the fact that the system does not currently address the significant occlusions, deformations or changes to the surface topology that often occur during the course of a typical liver resection task (all of which are still ongoing problems of interest in the research community).

In this system there are four notable subcontributions.

1. To the best of our knowledge, this system is the first to use an adaptive *Coarse-to-Fine (CTF)* stereo algorithm for fast and accurate 3D surface reconstruction in intraoperative liver imaging. The approach yields data-driven dense reconstruction by allowing coarse resolution image data to inform fine resolution reconstruction, even in low texture regions [179]. *Coarse-to-Fine (CTF)* processing also leads to computational efficiency, while complementary use of adaptive windows supports precise reconstruction of 3D boundaries.

2. We propose a shape-based (semi-) automated initialization strategy for registering an intraoperatively acquired 3D reconstruction to a preoperative *CT*

scan. To our knowledge this strategy is the first automated or semi-automated technique to provide quantitative results on intraoperative data and one of the first proposed for use in *Surgical Navigation (SN)*. Manual initialization strategies, that have become common practice in the *SN* community, have been shown to interrupt surgical workflow and the accuracy of the resulting registration is highly dependent on the expertise of the user performing the initial registration [14]. The proposed semi-automated strategy ameliorates this limitation.

3. The system makes use of a robust, 3D motion estimator based on interframe feature matching to register a time series of surface reconstructions to the initial frame of the sequence for subsequent registration to a volumetric *CT* scan. The proposed feature matching-based approach encourages registration to bring physically meaningful features into alignment and has shown to be resistant to drift in our experiments.

4. The creation of a controlled laboratory dataset allowing for ground-truth evaluation of the motion estimation ability of *SN* systems. To our knowledge this is the first laboratory controlled dataset mimicking open surgical settings with ground-truth signals for the motion of the phantom liver. One of the current challenges in the *SN* community is the standardization of evaluation of such systems. With this challenge in mind, the laboratory dataset has been shared with the *SN* community to enable future comparisons between systems.

The system has been evaluated (i) empirically in controlled laboratory experiments with a liver phantom for precise quantitative evaluation and (ii) using

patient data acquired during liver resections that took place at *Memorial Sloan Kettering Cancer Center (MSKCC)*. Both forms of evaluation take place in open liver resection (or analogue) conditions. Despite a shift in surgery toward less invasive procedures [191] and corresponding research effort in *SN* systems that target *Minimally Invasive Surgery (MIS)* applications, as evidenced by the focus of recent surveys in the area [96, 91, 128, 121, 113, 14], most liver resections are performed as open resections due to the extent and location of disease. A recent study from *MSKCC* reported that roughly 75% of liver resections at their centre were performed in an open environment [166]. Moreover, laparoscopic approaches are only suitable for certain cases where there is oncologic equivalency between open and laparoscopic interventions [166]. Overall, we demonstrate a stereo-based platform capable of providing temporally dense 3D textured data registration to a preoperative model in near real-time under realistic conditions of liver surgery. Portions of this work have been previously published [181].

1.4 Outline of dissertation

The dissertation consists of five chapters. Chapter 1 has motivated the work and outlined the contributions provided. Chapter 2 reviews related research. Chapter 3 details the developed technical approach to the problem of *SN* in liver surgery. Chapter 4 provides an empirical evaluation of the implemented system in laboratory and intraoperative conditions. Finally, Chapter 5 reviews the main points of the dissertation and suggests avenues for future work.

Chapter 2

Related research

Research on the development and use of *Surgical Navigation (SN)* systems is concerned with enabling the direct use of preoperative imagery during live surgery in a way that accounts for and is consistent with the intraoperative state of the patient. To date, developments in this field have required the knowledge and exploitation of various imaging modalities and subsequent manipulation of the chosen imaging modalities to provide common structures in the pre- and intraoperative images to place the two pieces of information into correspondence. Typically, *SN* systems of interest use preoperative volumetric imagery and attempt to extract similar 3D information intraoperatively (either surface estimates or additional volumetric imagery) and subsequently register the two pieces of 3D information to provide the common reference for the pre- and intraoperative states.

The contributions made in this dissertation consider algorithmic aspects of *SN* systems. Hence, in the literature review, we restrict discussion primarily to the algorithmic approaches documented in prior works. Little will be said about non-algorithmic details of the prior work (such as details of the surgical procedures,

Operating Room (OR) details, imaging setup, etc.). Furthermore, there are few datasets available and no attempt at common validation among published works. Metrics used for evaluation are also varied. Hard performance requirements and standardized metrics required to evaluate *SN* systems to determine clinical relevance is a pressing need in the community. Suggestions for accuracy requirements, as they pertain to abdominal and/or liver interventions, are provided for discussion within the empirical evaluation of the system presented in Chapter 4 of this dissertation.

A system classification based on functional pathways through extant *SN* systems is proposed and extant systems are reviewed. The system classification provides a way to distinguish between the main approaches to bridging pre- and intraoperative sources of imagery, while the extant systems provide specific examples of the various approaches yielding additional insight.

2.1 System classification and extant systems

This section provides a review of related approaches to bridging between preoperative volumetric imagery and intraoperatively acquired imagery. Much of the published research related to providing a bridge between preoperative volumetric imagery and live intraoperative imagery focuses on a subcomponent of the pipeline that is necessary to accomplish the task. Nevertheless, several fairly complete systems have been documented and are presented in this section. Figure 2.1 provides a framework for classifying the general approaches to this problem. The figure depicts the four main system designs that were identified in the current literature for relating intraoperative video data to preoperative volumetric data, as follows:

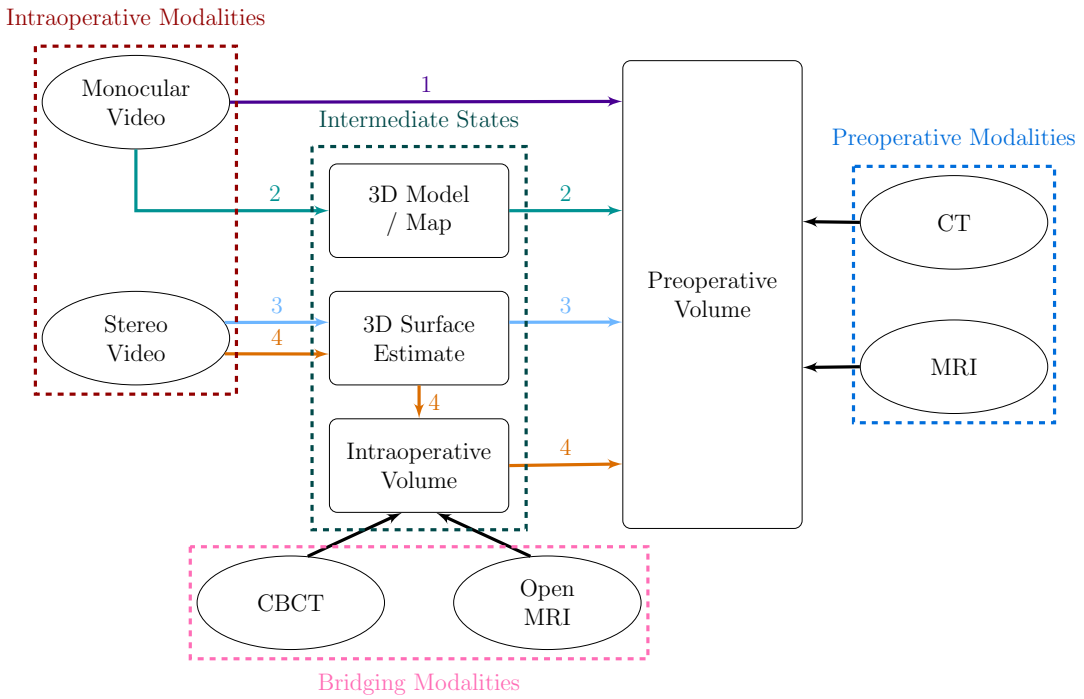


Figure 2.1: Functional Pathways Classification Through Extant Surgical Navigation (SN) Systems. Shown are the variety in approaches for systems relating preoperative volumetric images to intraoperative video imagery that appear in the current literature. Four primary paths exist through the classification of extant systems: **(1)** pose estimation from 2D-3D point correspondences, **(2)** pose estimation with respect to a model or map captured intraoperatively, **(3)** surface registration of a partial intraoperative 3D surface reconstruction and the preoperative volume, and **(4)** registration using an intraoperatively acquired volumetric image as a bridging modality.

- **Pose estimation from 2D-3D point correspondences.** **Path 1** in Figure 2.1 relies on a set of 2D-3D point correspondences relating visual features in a reference video frame to points on the preoperative volumetric surface. From this set of 2D-3D correspondences between the data sources, a projection can be estimated relating the two sources of data at a point in time. The projection can be maintained by tracking/matching the image features in subsequent frames of the video. While simple and relatively straightforward, this class of approaches is limited by the difficulty of detecting corresponding cross-modality features between the preoperative and intraoperatively acquired images.

- **Pose estimation with respect to a model/map captured intraoperatively.** **Path 2** in Figure 2.1 also employs monocular/2D video, but relies on a separate procedure that generates a surface map or model of the intraoperative state of the organ. This class of approaches attempts to bring the preoperatively acquired and intraoperatively acquired images into common representations (i.e., 3D surface models) and thereby overcome the challenges of establishing cross-modality point-wise feature correspondences. This process can be performed in a variety of ways (*Structure-from-Motion (SfM)* [207, 197, 74, 115, 123, 90], *Simultaneous Localization and Mapping (SLAM)* [130, 37, 56], *Shape-from-Shading (SfS)* [36, 116] as well as *SfS+SfM* hybrids [116] have all appeared in the literature). Once the intermediate 3D surface model has been computed from the video input, the intraoperatively constructed model can (optionally) be registered to the preoperative model using shape-based or point-based approaches (depending on the type

of map/model being maintained). Subsequent images can then be registered to the intraoperatively constructed model as it now contains visual features related to the 3D structure providing cross-modal features with which to match. Conversely, the ultimate ability for the system to localize the current intraoperative state to the preoperative imagery will be bounded by the cumulative errors of localizing to the intraoperative model, building the intraoperative model/map, and registering that map to the preoperative model.

- **Surface registration of a partial intraoperative 3D surface reconstruction and the preoperative volume.** Path 3 in Figure 2.1 makes use of 3D measurements at a given input frame (often via the computation of disparity from stereo imagery) along with shape-based registration techniques to register the shape of the viewed scene to the preoperative state of the organ. The use of stereo video (or another real-time 3D vision sensor) allows for the frame-by-frame 3D reconstruction of the visible scene. The ability to build a 3D surface estimate at every frame removes the necessity to build an intermediate 3D model/map based on the intraoperative data. As a result, any issues of distinguishing camera vs. scene motion are removed and scene motion can be recovered solely in the registration of the data sources. A main limitation of these techniques is in the large degree of local shape similarity typically present during the registration of small patches of the surface of organs and the preoperative imagery. This degree of local shape similarity can make the use of shape-based registration approaches particularly prone to finding local minima as well as having relatively "flat" solution spaces

where surface patches can be slid along the local region of the preoperative volume with relatively little change in the registration error [27, 188, 215]. Increasing the surface coverage of the intraoperative scan is currently the best non-invasive means to combat the issue of local shape similarity in these techniques [173, 158]. Similar approaches using *Laser Range Scanning (LRS)* or a tracked probe manually swept over the organ of interest have appeared in the literature (and in *Food and Drug Administration (FDA)*-approved systems in the case of tracked probes [175]). Recent work has explored miniaturizing *LRS* for endoscopic surgery by introducing two optically tracked probes: a single-plane laser as a structured light source and an endoscopic camera [55]. To reconstruct the visible anatomy in 3D, the endoscope is clamped in place while the laser is manually swept over the *Region Of Interest (ROI)*. These approaches have an important difference when compared to techniques using real-time 3D vision sensors. Though their pipelines may be similar, 3D reconstruction using *LRS* or a tracked probe is a time aggregated process. As such, systems relying on these 3D reconstruction techniques are only applicable for a "snapshot" approach to *SN* where results are shown at a single point in time as opposed to one that updates with the state of the patient. The novel system presented in this dissertation is in the class of approaches that make use of surface registration of a partial intraoperative 3D surface reconstruction to a preoperative volume.

- **Registration using an intraoperatively acquired volumetric image as a bridging modality.** Similarly to the previous option, **path 4** in Figure 2.1 makes use of an intraoperatively acquired volumetric scan (via

imaging techniques such as *Cone Beam Computed Tomography (CBCT)* or Open *Magnetic Resonance Imaging (MRI)*) as a bridging modality to aid in the registration process [6, 125]. These imaging modalities provide a means of obtaining a volumetric scan of the area of interest intraoperatively (typically at a lower quality/resolution when compared to their preoperative counterparts). Typically, the volumetric scan is used to bring additional context when compensating for deformations that may take place in the patient's anatomy between preoperative and intraoperative states. Partial surface models acquired at any time instant via stereo, or some other intraoperative 3D surface imaging modality, can be registered directly to the intraoperatively acquired volumetric scan. This intraoperative volumetric scan is in a similar physical state as the partial surface model. Utilization of the intraoperative volumetric scan allows the update registration process, that must occur frequently and with limited surface coverage, to deal with only a portion of the state change as a result, e.g., [125]. Meaningful integration of these scanners in *SN* systems is currently challenging due to issues including the large cost of such scanners, possible radiation exposure for the patient and surgical team, and the limited access for the surgical team to the patient while scanning.

The remainder of this chapter describes how extant systems in each of these categories operate.

2.1.1 Pose estimation from 2D-3D point correspondences

This general class of approaches uses hand-picked corresponding features in an intraoperative video frame and the preoperative 3D volume to constrain the estimation of the camera's pose with respect to the preoperative model [7, 187, 172, 142]. Figure 2.2 shows a typical configuration for a *SN* system in this category. These approaches assume that there are corresponding point features on the 3D surface and imagery that can be readily localized, which is often not naturally the case [138, 142]. With this in mind, the best that can be achieved in such approaches is a least squares approximation to the correct registration. A benefit of such an approach is that if you can assume a lack of any systemic bias in the error in identifying matching points the registration error will asymptotically improve as the number of point correspondences provided to the algorithm increase. However, if the additional points are not relatively uniformly distributed over the surface being registered, a small registration error can occur in the region with increased density of points at the expense of significantly larger errors in areas not well represented in the provided correspondences [118, 119, 49].

To avoid the dependency on hand-picked points, some systems make use of cross-modality fiducials that are placed on/in the patient's anatomy during a separate surgical procedure prior to acquiring the preoperative scans [7, 187, 172, 78, 192, 203, 77]. These cross-modality features can then be identified both in the intraoperative video imagery and the preoperative volumetric scan using automated or semi-automated means and then used to create a similar estimate of the relative projection relating the two sources of data, a process known as "inside-out" tracking. While these approaches simplify a portion of the technical challenges, they are

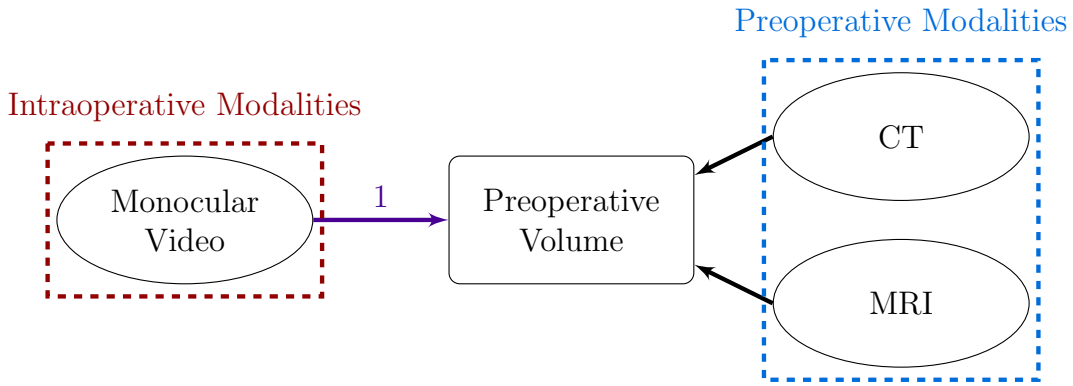


Figure 2.2: Pose Estimation from 2D-3D Point Correspondences. **Path 1** is extracted from Figure 2.1 for clarity of exposition and convenience of the reader.

often avoided due to the invasive nature of the additional surgical procedure needed to implant the fiducial markers [7].

Current approaches also typically assume that the intraoperative and preoperative states of the organ of interest have not changed in shape. These approaches typically end up computing the camera pose by estimating the projection that provides the best fit of the model feature points in the current frame, assuming rigidity. Once initial correspondences have been found in a reference frame, the tracking of the feature points in subsequent video frames allows for the projection matrix to be recomputed given the new locations of the feature points in the image frame [7, 187, 172, 123, 129, 142, 65, 77].

Exemplar system: Puerto-Souza et al. 2014

The system developed in this work aims to provide a long-term and accurate *Augmented Reality (AR)* surgical display during kidney *Laparoscopic Partial Nephrectomy (LPN)* surgery [142]. (Nephrectomy is a surgical procedure to remove all (radical) or part (partial) of a kidney.) The system attempts to directly compute a

projection matrix relating the 3D preoperative volume to the 2D intraoperative imagery.

To initialize the system, a series of 2D-3D point correspondences are needed to relate the preoperative volume to a single image frame in the intraoperative video. These point correspondences are obtained by projecting the preoperative model into the reference image frame and hand aligning the model to the image contour. Shi-Tomasi corner features [168] are extracted from the reference image frame with each feature being associated with the closest projected point from the preoperative model. Corners differing too far from the closest projected preoperative point are discarded from the initial set of "anchor-point" feature correspondences.

These 2D-3D point correspondences are then used to compute a 3×4 projection matrix that relates the preoperative volumetric 3D coordinate space to 2D pixel coordinates in the reference image. Different approaches to computing the projection matrix are used when the monocular camera is calibrated vs. when it is not. In the uncalibrated case, a version of the *Direct Linear Transformation (DLT)* [62] is used. When the camera has been calibrated, the images are first undistorted and then the *Perspective-Three-Point (P3P)* algorithm [97] is used. In both cases, a weighted-*Random Sample Consensus (RANSAC)* strategy [48] is used to make the projection estimation more robust. The weighted-*RANSAC* strategy works by lowering the contribution of clustered points, as well as points that had poor reprojection errors in the previous timestep, to the error computation used in a standard implementation of *RANSAC*. A *Sliding-Window (SW)* weighted least-squares approach [111] is used to help reduce jitter in the motion estimates.

To recompute the projection matrix relating the preoperative volumetric data to the intraoperative video, the 2D features that were extracted in the initial reference

image frame are tracked. Feature tracking is performed using an extension of the Shi-Tomasi tracker that accounts for affine transformations in the tracked template as well as changes in illumination and reflection [81]. The tracked features are then used to maintain the 2D-3D point correspondences between the preoperative volumetric image and the intraoperative imagery throughout the video sequence. When features are lost during tracking (i.e., due to sudden camera movement, occlusion, etc.) a tracking-recovery stage is deployed. In particular, *Scale-Invariant Feature Transform (SIFT)* features are extracted and matched across images before and after the failure [106]. It is argued that use of *SIFT* features and matching for tracking failure recovery is preferable to the simpler corner features used during regular tracking because they are more robust to scale and rotation changes. Finally, the initial correspondences are used in a *Hierarchical MultiAffine (HMA)* feature matching strategy [143] to compute an image transformation relating the position of the feature points in the image prior to the tracking failure to their location in the current image.

The system was evaluated on two monocular video sequences taken during *LPN* interventions. The proposed weighted *DLT* with *SW* approach was compared with three baseline approaches. First, a version of the system that estimates the projection matrix using *DLT* with *RANSAC* (i.e., without a weighted *RANSAC* and without *SW*). Second, a version of the system that estimates the projection matrix with the weighted *DLT* on a *SW* and the feature recovery stage. Third, a version of the system using a weighted *P3P* and *SW* and the feature recovery stage, requiring calibration information and rectified images as input. The video sequences contained a number of challenges for typical navigational systems including rapid camera motions, organ deformations, and prolonged occlusions. The first sequence is $\approx$19

seconds long, with most of the challenges resulting from changes in illumination and partial occlusions of the kidney. The second sequence is over 2 minutes in length and contains a wider range of challenging situations (greater tissue deformation, camera retraction/reinsertion, change in zoom, and total occlusions). The evaluation consisted of measuring the precision of the overlay, precision of the 3D registration, and the robustness to noise/occlusion. In general, the proposed approach provided a visibly more precise and stable overlay of the 3D model when compared to the organ boundary across the video sequences despite not always having the lowest reprojection and 3D errors across the various sequences. The average reprojection error ranged from 3.23 - 3.52 pixels across the three approaches for the first sequence and ranged from 3.39 - 3.59 pixels for the second sequence.

2.1.2 Pose estimation with respect to a model/map captured intraoperatively

As noted above, to overcome the typical lack of cross-modal features that are visible from both the intraoperative video and the preoperative volumetric model, some techniques perform a two-step reconstruction/registration approach [207, 37, 38, 56, 112, 90, 214]. Figure 2.3 shows a typical configuration for a *SN* system in this category. These architectures almost exclusively take advantage of properties of the tissues being registered and/or stages in the procedure that allow them to assume that the organ remains predominantly rigid while the reference model is being captured. Recent research has demonstrated 3D reconstruction of a scene from a monocular endoscopic video without relying on scene rigidity during the building of the intraoperative model [90]. This system uses a two-stream approach: *Shape-from-*

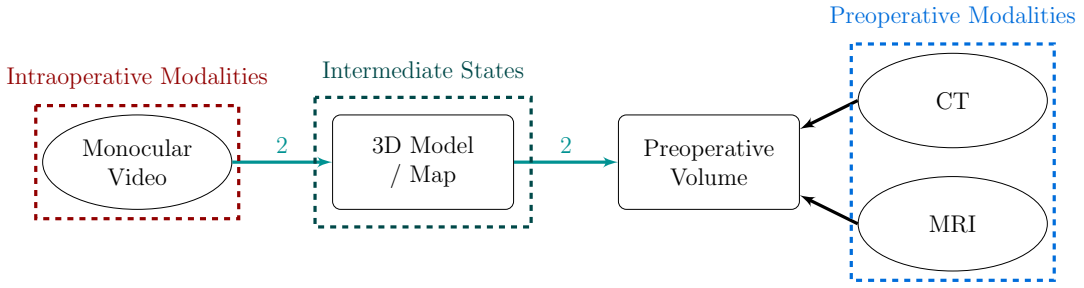


Figure 2.3: Pose Estimation With Respect to a Model/Map Captured Intraoperatively. **Path 2** is extracted from Figure 2.1 for clarity of exposition and convenience of the reader.

Template (SfT) is used to recover the camera pose and deformation information at frame rate with respect to the current state of the surface reconstruction and *Non-Rigid Structure-from-Motion (NRSfM)* is used to continually build and update the template map at a lower frame rate.

One version of this approach relies on the generation of a visually textured 3D surface map [37, 38, 90, 202]. The generated model provides a source of data relating the visible topology of the organ, in its intraoperative state, as well as visible image texture that can be seen by the cameras. Much research has been done on intraoperative surface reconstruction from monocular video in isolation (e.g., [206, 115, 186, 56, 90, 202]). To be able to relate intraoperative imagery with preoperative scans, the shape information contained in the model can be used to perform a shape-based registration between the intraoperative surface model and the preoperative volumetric model. Alternatively, registration between the intraoperative and preoperative information can make use of point correspondences between the imagery (or the map) and the preoperative model [207, 112].

When visual texture is encompassed in the map/model, image features extracted

in subsequent frames in the video can be used to localize with respect to the model through the use of image features contained in the intraoperative surface model, providing a registration chain back to the preoperative model [38]. Alternatively, visual tracking information [207] as well as external tracking (typically optical or magnetic tracking systems) [47, 151, 151] has been used to update the localization with respect to the model.

Some systems have been developed for cases where preoperative volumetric models are either not necessary for the surgical task [56, 110, 202] or are avoided to try and limit exposure of particularly sensitive tissues to radiation [176]. In cases where measurements and situational awareness require localization with respect to an intraoperatively acquired 3D reconstruction, registration back to a preoperative model may not be necessary. Two examples where *SN* techniques have been applied to a task requiring only intraoperative localization include measuring the size of a hernia repair [56] and diagnosis/localization of lesions inside the stomach during gastric endoscopy [202]. Alternatively, in the field of sinus surgery, it is common practice for the surgery to take place without patient specific *Computed Tomography (CT)* information unless absolutely necessary [176]. An approach to *SN* in this domain uses statistical shape models, derived from the *CT* scans of previous patients, to act as a generic preoperative anatomical model [176]. A deformable registration is performed between the 3D surface estimates generated during the endoscopic video sequence and the statistical shape model. The deformable registration along with the shape model is used to provide a visualization of the estimated surrounding structures, along with an associated confidence, as a coarse localization aid to the surgeon.

Exemplar system: Zhou et al. 2019

A system for intraoperative reconstruction of various tissue surfaces from stereo laparoscopic videos is presented by Zhou et al. [215]. The system uses a feature-based *SLAM* algorithm to mosaic 3D surface estimates from individual stereo frames (captured at a close standoff distance) to generate a coherent 3D model of the viewed tissues.

The stereo matching component of the system was selected to avoid both the complexity of global stereo matching methods while allowing the system to provide stable and dense matches in low texture regions (something typically difficult for local stereo matching methods). The system uses a local block-matching approach with a *Zero-Mean Normalized Cross-Correlation (ZNCC)* similarity metric to determine dense correspondences across the stereo pair. The algorithm uses 11×11 windows, analyzing every other pixel in a checkerboard pattern. Given a pair of 960×540 images and a disparity search range of 100 pixels the runtime on a *Graphics Processing Unit (GPU)* is ≈ 5 ms.

The matches from the stereo correspondence method are typically quite noisy and include a large number of outliers. To provide a cleaner 3D reconstruction, the system runs three postprocessing steps: outlier detection/removal, hole filling, and surface smoothing. Outlier detection starts off by assuming that the tissue surface is relatively smooth, throwing out any point that does not have a sufficiently close disparity value in one of its neighbouring pixels. Two hole filling methods are performed when a match cannot be found between the two stereo views at a pixel location. The first approach searches each of the eight radial directions within a 50 pixel range, while the second approach searches all pixels around

the location in need of a disparity value that are within a 20 pixel radius. For both approaches, if a sufficient number of points in the search range have a valid disparity value, the missing value is filled via interpolation. These two steps are performed in an iterative process (alternating between outlier removal and hole filling) while gradually increasing the outlier rejection radius. This allows disparities that are removed during outlier rejection to correspondingly be filled again gradually providing an increasingly smooth surface. The last step for processing the individual frame 3D reconstruction is to perform a modified Laplacian smoothing procedure [195].

Point based registration is then performed on adjacent frames using extracted *Oriented Features from Accelerated Segment Test and Rotated Binary Robust Independent Elementary Features (ORB)* features [157]. Given the viewing conditions typical in laparoscopic surgeries, outlier rates on matched features were seen to be higher than 85% on occasion. In these cases, the traditional *RANSAC+P3P*-based outlier rejection method did not perform well and was prohibitively slow [88]. As such, the authors leveraged an infinitesimal camera motion assumption and used the displacements of matched *ORB* features as an outlier criteria.

The DynamicR1PnP algorithm is used to estimate the camera position and orientation from n feature matches between the 3D reconstruction (available at the 2D feature locations of the previously reconstructed frame being registered to) and their corresponding 2D image projections (in the frame being registered) [215]. The method incorporates a re-weighting strategy and the *1-Point Random Sample Consensus (1PR)* algorithm to reduce the effects of outliers.

Keyframes are used in the system to help limit the frequency of having to calculate the global model. A keyframe is added if it satisfies the following

conditions: there are at least 50 correct *ORB* matches after the DynamicR1PnP stage and it has been at least 10 frames since the last keyframe or the camera pose has changed greater than some threshold since the last keyframe.

Given the camera motion tracking results, when adding a new keyframe to the map, keyframes with significant overlap to the current keyframe are selected as candidate keyframes. *ORB* features at the current keyframe are matched to features of the candidate keyframes and DynamicR1PnP is run to help eliminate some of the accumulative error. A cost function combining local *Bundle Adjustment (BA)* costs (to minimize reprojection error) and *Iterative Closest Point (ICP)* costs (to minimize the distance between the global 3D model and the current stereo matching reconstruction) is minimized to produce the updated camera pose estimates.

Once aligned, the *Truncated Signed Distance Function (TSDF)* method is used to mosaic the raw 3D point clouds to obtain an extended 3D model of the viewed tissue surface. This process is essentially averaging and merging the 3D coordinates if multiple observations lie within some area about that point. Pixels with a valid disparity measure that do not have a match in the extended surface model are considered as new observations and are simply added to the extended surface model. To texture map the model, given that the light source is located at the tip of the laparoscope, pixels at the center of the image are updated with a weight of 1 with the weight decreasing towards the edge of the image during a *TSDF* merge.

The system was tested on ex and in vivo data. The ex vivo data was conducted using four porcine livers and three porcine kidneys. Registration between the intraoperative surface model and the preoperative volumetric model was performed by segmenting the preoperative *CT*, manually selecting landmarks such as tissue tips, edge points and other recognizable landmarks, and then refining the registration

with *ICP* over the entire 3D surface reconstructions. *Root Mean Square Error (RMSE)* errors on the four liver cases ranged from 1.1 - 2.0 mm and ranged from 1.0 - 1.1 mm in the three kidney cases.

The majority of the leveraged algorithms had been parallelized and optimized for computation on a *GPU*. The average runtime per keyframe was reported as 76.3 ms on stereo images with 960×540 resolution. The authors note that the current pipeline relies on the assumption that the scene remains static and thus should be applied only to surgical scenarios with minimal deformation present. Under these conditions the authors state that the reconstructed surface could be used for further registration to an intraoperatively acquired volumetric image.

2.1.3 Surface registration of a partial intraoperative 3D surface reconstruction and the preoperative volume

Perhaps the most prevalent class of systems in the literature relies predominantly on shape-based registration for relating intraoperative imagery to preoperative volumetric images. In these systems, relating the 2D imagery with the preoperative volumetric imagery first requires the extraction of 3D shape information from the video source. Figure 2.4 shows a typical configuration for a *SN* system in this category. When stereo imagery is used as the input modality this operation is achieved by first computing the corresponding locations of points across the pair of stereo images. Subsequent triangulation of the matched points (correspondences) into 3D space results in a 3D point cloud capturing the 3D structure at these locations for each input stereo frame. Considerable work has focused on intraoperative surface reconstruction as an independent research contribution (e.g., [94, 12, 28, 183, 108]),

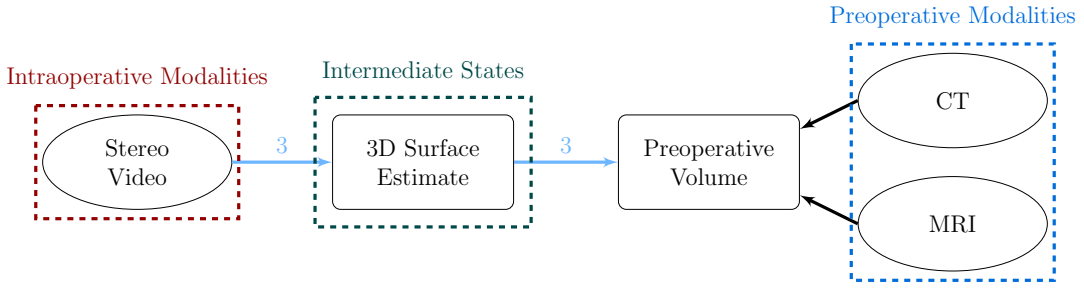


Figure 2.4: Surface Registration of a Partial Intraoperative 3D Surface Reconstruction and the Preoperative Volume. **Path 3** is extracted from Figure 2.1 for clarity of exposition and convenience of the reader.

often with a stated desire to extend these contributions in the future so that they register the reconstructions to preoperative imagery for *SN* assistance.

The recovered shape of the 3D surface can be registered to the preoperative volume (typically with some manual assistance to begin) using shape-based registration techniques to relate the stereo imagery to the preoperative model [42, 193, 185, 151, 152, 153, 60, 188]. Notably, a recent survey noted that, to their knowledge, all shape-based intraoperative registration methods, used in laparoscopic interventions, relied on *ICP* or one of its variants [113].

The majority of systems that fall into this category make use of stereo vision techniques to generate the intraoperative estimate of the 3D structure of the surface of the organ. Some techniques extract primitive features such as central lines of the coronary tree [42] or salient landmark/point-based features chosen manually [185] or automatically [60]. When paired with epipolar constraints, meaningful (albeit sparse) 3D surface information can be extracted for the organ, typically via a simple nearest neighbour strategy [60]. Other approaches often try to provide a more dense estimate of the 3D structure of the scene [193, 151, 152,

153, 188, 211, 214]. While many approaches to dense stereo estimation exist in the medical imaging literature [113], only a few examples of integration into a system attempting to relate recovered shape back to preoperative imagery is evident in the literature. One solution attempts to leverage a standard *Dynamic Programming (DP)* based semi-global approach to recovering 3D structure of the kidney [193]. Ordering constraints are used, in a *DP* paradigm, to provide an optimized solution (under the leveraged constraints) on a row-by-row basis in the rectified stereo pair. Another solution, using a semi-global technique, *Hybrid Recursive Matching (HRM)* originally developed for teleconferencing applications [4], has been applied to *SN* applications targeted at the heart and liver [151, 152, 153]. The *HRM* algorithm uses the previously processed nearby spatial locations and a temporal disparity candidate at the same image location in the previously processed image frame to "bypass" the local search process from traditional search methods due to the large search range present in their application. Limiting the search to so few candidates is unlikely to provide the correct answer on its own. As such, it is paired with a non-standard meandering scan path (to provide optimal spacing of spatial disparity candidates) and a gradient-based iterative update technique (to refine the initial disparity candidates) to provide a dense disparity map ready for post processing that enforces boundary-aware smoothness constraints.

Exemplar system: Röhl et al. 2011

A system to reconstruct the surgical surface from stereo endoscopic images is presented by Röhl et al. [151, 152, 153]. The system was designed for use in real-time with millimetre accuracy for intended intraoperative use in an *AR* setup [151]. To assure the quality of the stereo reconstruction, the stereo endoscope is calibrated

prior to every intervention. As a preprocessing step, each stereo pair is rectified from this preoperative calibration. Additionally, simple specular highlight detection is performed using an intensity threshold.

To generate dense disparity measurements from the stereo pair, a semi-global approach is used, *Hybrid Recursive Matching (HRM)* [4]. The core aspect of this algorithm is to minimize the computational effort of disparity search by reducing the search to a small number of candidate disparities generated from the spatial and temporal neighbouring disparity estimates at each pixel.

Several modifications to the generic system were made to apply this algorithm to the medical domain. The system made use of an alternative error metric (simplified census [151]) to help cope with the surface reflectance of the tissues being viewed. They also extend the algorithm to calculate subpixel disparities as the original design of the algorithm was meant for large-baseline stereo applications. The ordering of pixels in the disparity map, along with the removal of points not obeying the ordering constraint, is exploited to perform a simple meshing on the point cloud. For registration, an external tracking system is used to track the endoscope. A chessboard pattern is used to generate the transformation between the endoscope tip and the external tracker coordinate system. The transformation is calculated by using a tracked pointer device to obtain the locations of the chessboard corners in the external tracking coordinate system and registering these points to the reconstructed 3D positions of the chessboard corners from the stereo imagery. The endoscope can then be tracked in the external tracking coordinate system. Finally, the reconstructed surface requires an initial registration to the preoperative model, which is done by hand. *Iterative Closest Point (ICP)* [29, 15] is performed to help refine any error present in the surface reconstruction/endoscope

tracking pipeline.

An extension that attempted to use bilateral filtering [136] (and joint-bilateral filtering [117]) techniques for disparity smoothing and depth propagation in semi-occluded regions was later explored [152]. The work was also extended to a joint *Central Processing Unit (CPU)-Graphics Processing Unit (GPU)* implementation that showed improved accuracy of depth measurements and faster processing times (due to repeated computation of the disparity maps: low resolution results on the *CPU* and high resolution, tiled results on the *GPU*) [153].

The system was evaluated against an OpenCV [20] implementation of semi-global block matching [72] as well as a constant-space belief-propagation algorithm [71] (both employed with default parameters from the OpenCV library with no further tuning to the particular dataset). A custom simulator was made to help evaluate the algorithms against different camera motions (allowing for more or less temporal support for the proposed *HRM* method), various textures (ranging from no-texture to several textures based on laparoscopic images of liver tissue), image noise, and errors in rectification. All methods for computing disparity were able to achieve subpixel mean disparity error and sub-millimetre reconstruction error across all simulated (non-noisy) datasets with the *CPU-GPU* hybrid approach edging out the competition across most test configurations. With the introduction of image noise and rectification error the *CPU-GPU* hybrid approach and the semi-global block matching were more evenly matched.

Additional tests were performed using image sequences (and corresponding ground-truth point clouds from manually registered *CT* scans) of a heart phantom. Again, the *CPU-GPU* implementation was evenly matched with the semi-global block matching approach when comparing the mean error in registration distance

across all of the phantom-based datasets.

Qualitative analysis was performed offline on in vivo data collected during a laparoscopic surgery performed by a da Vinci surgical robot[1] [79]. The authors noted that the results from the semi-global block matching algorithm were visibly noisier than their proposed algorithm in the presence of surgical smoke and had problems reconstructing very homogeneous regions. They noted this result was likely due to the temporal support present in the *HRM* as well as its recursive implementation. The two algorithms also worked equally well in the presence of specularities, but the semi-global block matching algorithm performed worse in undersaturated/dark regions of the image. The semi-global block matching algorithm performed better in regions with depth discontinuities (e.g., near the presence of surgical tools).

The authors noted that the *GPU*-only version of the system performed fast enough to be used intraoperatively (achieving 40 *FPS* using an image resolution of 320 $\times$ 240 pixels) [151]. Their recommended *CPU-GPU* method achieved a processing rate of 14 *FPS* using an image resolution of 640 $\times$ 480 pixels in a follow-on publication [153]. The 3D reconstruction portion of the pipeline was evaluated in isolation (with known registration transformation parameters) using a computer generated liver simulation as well as with a silicone heart phantom. The recommended *CPU-GPU* method achieved a mean distance error of 0.08 - 0.56 mm in the sequences generated in the liver simulation and 1.45 - 1.64 mm in the phantom heart datasets. The quality of the surface reconstruction along with initial overlay of the pre- and intraoperative models was deemed sufficient to successfully register the two models.

[1]The da Vinci Surgical System is product of Intuitive Surgical, Sunnyvale, CA, United States.

2.1.4 Registration using an intraoperatively acquired volumetric image as a bridging modality

The last set of systems rely on the recent introduction of intraoperative volumetric scanners present in modern hybrid operating rooms. Figure 2.5 shows a typical configuration for a *SN* system in this category. *Cone Beam Computed Tomography (CBCT)* and Open *Magnetic Resonance Imaging (MRI)* scanners are able to provide volumetric images of the patient's anatomy while they are present on the operating table. Meaningful integration of these scanners in *SN* systems is currently challenging due to several issues, as follows. First, costs associated with the equipment and its utilization are high [173, 14]. Second, there is often a need to modify the surgical environment. Open *MRI* sensors require the removal of ferromagnetic materials to help avoid magnetic interference [69, 173]. All intraoperative volumetric scanners take up non-negligible space in the operating theater and can easily necessitate modifications to the typical surgical workflow due to issues relating to access to the patient [173]. Third, additional concerns of radiation exposure to the patient and surgical team are present with the use of some intraoperative volumetric scanners, including *CBCT* scanners. Fourth, the time required to acquire and postprocess intraoperative volumetric images is non-negligible in comparison to the targeted use cases of *SN* systems [69, 14].

Nevertheless, intraoperative volumetric scanning has been used in select research approaches to the problem of relating intraoperative video views to preoperative volumetric imagery [122, 125, 13]. In these techniques, the intraoperative volumetric scan is used as a bridging modality between the intraoperative video and the preoperative scans. The additional information present in the intraoperative

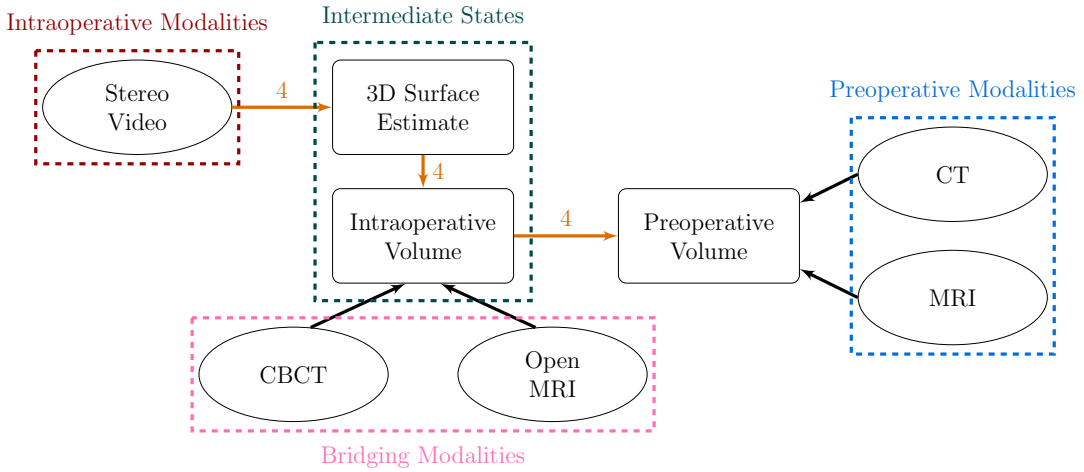


Figure 2.5: Registration Using an Intraoperatively Acquired Volumetric Image as a Bridging Modality. **Path 4** is extracted from Figure 2.1 for clarity of exposition and convenience of the reader.

volumetric images allows for more aggressive registration techniques for relating the intraoperative and preoperative states of the patient's anatomy while doing so in a more data-driven manner than is typically possible with only a partial surface reconstruction of the patient's anatomy and few (if any) cross-modality features automatically available for the system to key in on.

Exemplar system: Mountney et al. 2014

A comprehensive system for augmented reality in soft-tissue *Minimally Invasive Surgery (MIS)* is provided by Mountney et al. [125]. The key contribution of this work is to introduce a *Cone Beam Computed Tomography (CBCT)* imaging modality into the system pipeline that produces 3D *CT*-like images and 2D fluoroscopy in the same coordinate system while the patient is on the operating table.

When used as a bridging modality, this approach allows for automated co-registration of preoperative *CT* information and laparoscopic images. This proce-

dure is accomplished in three stages. First, a registration of *CT* to *CBCT* that takes into account large surface deformations due to the insufflation process (gas being introduced into the abdomen to make space for surgical tools). Second, a registration of the laparoscope to the *CT* via the *CBCT* coordinate system accounts for minor residual offsets and local deformations due to respiration and pulmonary activity. Third, a temporal registration across laparoscopic images accounts for camera motion and tissue deformation caused by respiration and pulmonary activity.

Two preoperative scans are performed using a contrast injection at the arterial and venous phases. These *CT* scans are registered together and segmented into 3D anatomical models such as the liver, tumour, vessels, abdominal wall, etc. During the procedure the abdomen is insufflated with CO_2 causing a large initial deformation to occur. The *CT* is registered to the *CBCT* coordinate system using a non-rigid biomechanically based registration technique [132]. The general approach is to perform a rigid alignment of the spine, model and account for the insufflation process (making use of the abdominal pressure being measured), and then perform a diffeomorphic non-rigid registration. The final deformation field can then be applied to preoperative data to bring it into the current, *CBCT* aligned, state of the organs.

Registration of the laparoscope to the *CBCT* coordinate system is done in a two step process: an initial position estimation followed by local refinement. The initial position of the laparoscope is estimated using fluoroscopic images. The laparoscope is held stationary by a mechanical device and two mono fluoroscopic images are acquired (each separated by 90°). A semi-automated approach selects two points along the shaft of the laparoscope, which are then triangulated to

estimate five-*Degrees of Freedom (DOF)* of the laparoscope's pose. The rotation about the camera axis and the position of the camera in the shaft are not recovered using this approach, which introduces a source of error down the pipeline.

A local registration refinement is then performed directly between the laparoscopic images and the *CBCT*-aligned 3D surface model of the organ. At this stage, the patient is not at breath hold; therefore, it is important to again perform the registration when the patient is at a fully inhaled state to prevent introducing further sources of error. (Breath hold is a technique that can be used to mechanically stop the breathing of a patient on a ventilator [70]. It can be used to temporarily cease the motion in the surgical field that the patient's respiration induces in a safe and controlled manner.) To help ensure this result, features are tracked on the tissue surface, matched in the left and right images, transformed into the *CBCT* space (to filter out points not showing as being on the tissue surface), and then subjected to *Principal Component Analysis (PCA)* and smoothed to recover a 1D respiration signal for each feature. After performing outlier rejection, the localized respiration signals are averaged to produce a global respiration signal. Parameters of the respiration model are estimated for the global respiration signal using the *Levenberg-Marquardt (LM)* minimization algorithm and the maximum inhale state is extracted. A point cloud is extracted from the preoperative model (using the laparoscope's position, the camera's intrinsic parameters, and z-buffering) and is registered to the reconstructed point cloud (generated by propagating disparity values about matched features [183]) using a probabilistic version of *ICP*. The use of a probabilistic *ICP* approach helps compensate for the local differences in shape due to the different approaches used to generate the point clouds.

The features being matched are generated as a result of a learning algorithm [126].

The feature tracker attempts to learn what information is most suitable for tracking in the *MIS* environment while the system is online.

Finally, the laparoscopic images are temporally registered. The tissue deformation and laparoscope position are jointly estimated in a modified, *Motion Compensated, Simultaneous Localization and Mapping (MC-SLAM)* framework [127]. The *SLAM* framework uses an *Extended Kalman Filter (EKF)*. The camera's state (position, orientation, translational velocity, and angular velocity), respiration model parameters, and the map are encoded in the state vector. The camera motion is predicted using a constant translational and angular velocity model. The state prediction model predicts the point in the respiration cycle and subsequently predicts the motion in the map points. The measurement model transforms state space/world coordinates into measurement space/image coordinates. Features are then tracked using the tracker mentioned above [126]. To initialize the system, the camera is assumed stationary for one respiration cycle.

The system was evaluated on simulated data (mesh from a *CT* scan textured with laparoscopic images), silicon liver phantom data (with surface fiducials for ground-truth), offline ex vivo porcine data (with fiducials), offline in vivo porcine data (without fiducials), and non-medical synthetic data from the Stanford dataset again textured with laparoscopic images. For every dataset that was captured, corrupted versions of the dataset were created with random noise added to the ground-truth starting position of the laparoscope in the *CBCT* system. Two error metrics were published: *Surface Registration Error (SRE)*, which captured the average metric distance between each point in the registered point cloud to the closest preoperative surface point, and *Target Registration Error (TRE)*, showing the *Root Mean Square (RMS)* error of the fiducials between their known ground-

truth values and their position after registration. The system was able to provide approximately 1 mm accurate *SRE* results (ranging between 0.8 mm and 1.3 mm across datasets) while experiencing significantly larger *TRE* results (0.3 - 1.69 mm in simulated datasets and 3.4 - 4.1 mm in ex/in vivo datasets).

2.1.5 Learning-based systems

Much like the majority of sub-disciplines in computer science (especially research areas in computer vision), *Machine Learning (ML)* has found increased adoption into recent *SN* systems. Current integration of *ML* techniques into *SN* systems generally follows the system classification presented earlier in Section 2.1 and replaces one or more components in the pipeline with a learned component. Pertinent to *SN* systems, independent contributions towards learning-based components have been gaining prominence in recent years. Contributions in areas such as *Deformable Image Registration (DIR)* [89, 150, 180, 209, 39, 46, 76, 167, 5, 16, 40, 67, 68, 75, 95], image segmentation (of volumetric [80, 213] and 2D imagery [154, 109, 64]), structure recovery from monocular [103, 216] and stereo images [21, 210], and activity recognition [217] have all appeared in recent years in the literature. Recent reviews have been produced to summarize the efforts in these areas [101, 30, 124, 63, 107].

While many of these self-contained research areas have started to have an influx of *ML* techniques being applied to the tasks at hand, integration into more complete *SN* pipelines has been slower. Some research into *DIR* is targeted at *SN*-like applications in the context of hybrid *ORs*, where the current state of the tissues could potentially be captured by consistent/regular volumetric imaging (e.g., [99, 40, 67, 68]). This approach, while attractive from a research perspective,

appears to currently produce several barriers to surgical workflow (discussed in Section 2.1.4). Integration of *ML* components into more traditional *SN* pipelines, as depicted in Figure 2.1, has occurred by replacing one or more components in the pipeline with a learned version of the same component (e.g., [104, 110, 202, 211]).

Deep Neural Network (DNN)-based depth estimation networks have seen integration into some *SN* systems (e.g., [104, 110]). The single-frame depth estimation network in both of these *SN* systems is based on an encoder-decoder architecture with skip connections between the encoder and decoder layers at the same scale. One system design interleaves recurrent units into the encoder to integrate spatial and temporal information in the depth estimation process [110, 200]. The same *SN* system supplies the current image along with the predicted depth map to a second network, based on the *Visual Geometry Group (VGG)*-16 architecture [171] with recurrent units interleaved, for visual odometry estimation [110].

In a stereo *Simultaneous Localization and Mapping (SLAM)*-based *SN* system for *Laparoscopic Partial Nephrectomy (LPN)* surgery, a Mask *Region-based Convolutional Neural Network (R-CNN)* [66] has been used to segment the portion of the 2D image that belongs to the kidney [211]. The segmentation allows for 3D reconstruction to be restricted to the kidney, in isolation from surrounding tissues/objects. Here, a *Region Proposal Network (RPN)* proposes candidate object bounding boxes, a feature extractor acting on each candidate box is used for classification and a per pixel mask is generated by a *Fully Convolutional Network (FCN)* [105] for each *Region Of Interest (ROI)*. *DNNs* have also been used as a feature extractor. In one *SN* system, this technique allows for images of the same *ROI* to be identified in an interventional video sequence and combined to produce a high resolution surface reconstruction of that *ROI* using *Structure-from-Motion*

(SfM) methods [202].

Exemplar system: Zhang et al. 2019

One example of integrating a learned component into an *SN* system was targeted at providing video see-through *AR* navigation for *LPN* (kidney) surgery [211]. The system relies on stereo laparoscopic video to provide depth estimates of the visible surface of the patient's kidney. Learning in this system is used to segment portions of the images that correspond to the patient's kidney and restrict subsequent point clouds to arise from these regions. These individual surface estimates are stitched together to provide a global surface estimate that is registered to the preoperative model of the patient. In regards to the system classification already presented, this system follows the pattern of pose estimation with respect to a model/map captured intraoperatively, Section 2.1.2.

Dense stereo matching is performed via *Semi-Global Block Matching (SGBM)* [71, 72]. The particular semi-global approach used in this system makes use of local/pixel-based matches using a *Sum of Absolute Differences (SAD)* sub-pixel match metric [17] while approximating a global smoothness optimization via a series of 1D regularization calculations to help promote smooth changes in disparity in multiple directions. *SGBM* applies regularization across multiple directions in an attempt to avoid the types of streaking artefacts that can occur when regularization is restricted to individual scanlines (such as in typical *DP* stereo matching approaches).

Renal surface segmentation in the laparoscopic image is performed using Mask *Region-based Convolutional Neural Network (R-CNN)* [66]. Mask *R-CNN* is an

extension of the Faster *R-CNN* [145] architecture that adds a *Fully Convolutional Network (FCN)*-based branch [105] for predicting pixel-wise segmentation masks on each *ROI*. Transfer learning is leveraged during training of the network. The network is pre-trained on the MS COCO object recognition dataset [100] and refined using a task-specific dataset consisting of manually segmented renal regions on a set of 1062 clinical laparoscopic images from nine different patients. The trained network is applied to an incoming laparoscopic image and the kidney is segmented from the rest of the surgical scene in the form of a mask output from the network. The mask is then used to restrict the 3D reconstruction at that frame to the region of the image containing the kidney surface.

In an attempt to decrease the ambiguity present when registering the frame-wise intraoperative reconstructions with limited *Field-of-View (FoV)* to the preoperative volumetric model, individual reconstructions of each stereo frame are stitched together to produce a global intraoperative surface model of the kidney. The stitching process begins by detecting and matching *Speeded Up Robust Features (SURF)* features in images taken at sequential timesteps [8]. *RANSAC* is then used to eliminate outlier matches [48]. The 3D locations of the matched 2D features are generated from their corresponding disparity maps and a point-based registration method is used to generate the 3D transformation relating the two stereo frames. The 3D reconstruction of the incoming frame is warped back to the initial frame, using the chain rule to extend the transformation back to a keyframe. Points that do not overlap (after transformation) are added to the global reconstruction to grow the surface map. Global bundle adjustment is then performed to reduce accumulated errors across the sequence. The resulting point cloud is stitched and downsampled using a box grid filter to reduce noise.

After generating the global intraoperative surface reconstruction of the kidney the final task is to register the intraoperative reconstruction to the preoperative surface model. As part of the preprocessing, the half side of the kidney that will be exposed to the laparoscope is manually selected from the preoperative model. *PCA* is then used to produce an initial coarse alignment [83, 199]. With an initial alignment in place, *ICP* is then used for rigid refinement [15]. Finally, *Coherent Point Drift (CPD)* is used for a final non-rigid refinement.

The system was evaluated on a kidney phantom in laboratory settings as well as an in vivo clinical stereo video sequence. The rubber kidney phantom had 14 pins with an average diameter of 1.7 mm inserted into the simulated organ to act as cross-modality features. A *CT* scan of the phantom (including fiducials) was captured before the evaluation. Real kidney surface texture was projected onto the phantom using an overhead projector and the phantom was placed in an abdominal model. Reconstruction error on the phantom (as measured as closest Euclidean distance to the preoperative model) was typically ± 1 mm. Tests were then performed where an instrument placed deformation by a point source on the rubber phantom (where deformation at that point ranged from 5 - 10 mm) and reported that the mean *TRE* was 1.28 mm (with a maximum *TRE* of 2.60 mm) under such conditions. In vivo tests were performed offline and were qualitative in nature (showing the overlay of the preoperative *CT* scan in the laparoscopic image).

Results of the *CNN*-based segmentation method was also reported in isolation. Predicted segmentation masks of never before seen laparoscopic images were compared with manual segmentation using an *Intersection over Union (IoU)* metric. Images taken from 15 different capture angles were used to verify the

system. The mean segmentation accuracy rate was 94.9% with a segmentation time of 0.6 seconds per image.

The authors noted that the system still had three main conditions to be met for the system to perform adequately. The first constraint was that the global reconstructed surface model needed at least 30% coverage of the preoperative model for an accurate initial registration between the intraoperative and preoperative models to be achieved. Second, the deformation of the target organ had to be minimal to ensure that the stitching process would represent patches taken of the organ in the same deformation state. Third, as this model is built over time, to ensure the successful implementation of the navigation framework, the target organs need to remain relatively stationary during stitching.

2.1.6 Hybrid systems

Some systems have appeared in the literature that have implemented several paths in the system classification [25, 193, 7]. Presumably, such combinations allow for systems to compensate for when the system is faced with not enough information to perform in its preferred manner by maintaining the registration between intraoperative and preoperative datasets through the use of a secondary method.

Exemplar system: Vagvolgyi et al. 2008

An interesting system for augmented reality in robotically assisted *Laparoscopic Partial Nephrectomy (LPN)* surgery was produced by Vagvolgyi et al. [193]. Here, determining the precise border between tumour and normal kidney tissue can be

challenging despite the use of intraoperative ultrasonography. The introduction of *AR* technology to this task can help accurately identify tumours (and their margins) and help adoption of minimally invasive surgical techniques for partial nephrectomy.

Prior to the operation, the preoperative *CT* image is segmented manually to generate a 3D surface model of the kidney, tumour (or stone), and collecting system. During the surgery, as part of the system initialization, a segment of the recorded stereoscopic video with the kidney in view is used to manually register to the segmented 3D-*CT* kidney model. To accomplish the manual registration, the segmented kidney model is overlaid on the endoscopic video segment and manually adjusted to obtain the best visual fit. After this procedure, surface-based tracking techniques are used to further refine the manual registration. The tracking system makes use of several manually selected corresponding points, in both the *CT* scan and the imagery, on the kidney surface surrounding the renal mass as fixed reference points. Further refinement of the registration is done via the *Iterative Closest Point (ICP)* algorithm [29, 15].

The system generates dense disparity estimates using a custom dynamic programming solution [193]. When shape is distinctive enough, it can perform a multi-stage 3D-3D registration. The rigid transformation between the reconstructed surface and the preoperative *CT* model is estimated using *ICP*, then a deformable surface registration is computed using a spring-mass system. When shape is not enough, the system makes use of sparse registration methods. For this process to work an initial registration between the video and the *CT*-model has to be achieved (either via a successful dense registration or manually). Once a registration is known, the system relies on a set of manually selected and tracked 2D features to constantly

update the registration between frames [193, 185].

The system was evaluated qualitatively offline on in vivo data collected during an animal trial as well as on two in vivo patient datasets. For the animal trial, the stereo algorithm was able to cope with the discontinuities present in the anatomical structure, but had some difficulties recovering depth values around the surgical tools. No preoperative *CT* data was available for registration results so limited tests were performed based on an initial clean frame being used to test the registration. In comparison to the reconstruction of the clean frame, the average error per vertex after rigid registration was below 2 mm with deformable registration techniques reducing the average error to below 0.5 mm for most of the same video segment. The first patient dataset was captured during a *LPN*. The system was able to accurately overlay the cutting margins on the kidney surface around the tumour in an offline process. The accuracy of the overlay was verified by the surgeon who performed the surgery. For registration purposes, an initial registration was performed by hand with sequential updates using pinned surface feature points. The second clinical case covered the removal of a large kidney stone. A manually segmented *CT* scan was used to show the kidney surface as well as the kidney stone. The same registration and pinning process was used to update across a video sequence that was approximately 1 minute in length. No empirical evaluation was performed on the in-vivo patient data.

2.2 Open issues

Despite the significant increase in activity of research into vision-based structure estimation and registration of intraoperative imagery to preoperative images in the

SN community, such systems are still not a part of the typical surgical workflow for many surgical disciplines. Nevertheless, the promise of such technology is an appealing one. Indeed, being able to provide an up to date alignment between preoperative imagery and the current state of the operating environment will open up new possibilities during surgery. This alignment will allow for accurate and patient-specific navigation of instruments and the ability to execute preplanned manipulation of organs [14]. Similarly, estimated 3D structure has potential to provide information that enables robot surgical assistants to perform with greater autonomy than currently is the case [91]. Eventually, these new abilities may open up new possibilities for the types of ailments that may be treated and significantly change the severity of some treatments that currently have non-negligible risks for complication [169]. Indeed, relevant and necessary steps are being taken in the *SN* community to make this technology a reality, but there is still much ground to cover for most of the intended applications.

2.2.1 Optical organ registration

Existing systems for optical organ registration in surgery have been proposed; however, they typically rely on techniques that may ultimately hinder their adoption in soft tissue surgery. Point-based [7, 187, 172, 142] and surface-based approaches [42, 34, 151, 152, 153, 60] are dominant for initializing the registration between preoperative and intraoperative scans; however, these approaches are currently hindered by being heavily manual. Volume-based approaches can be used to make this initial registration more automated and robust [125, 13], but are hindered by their underlying technology: imposing additional workflow concerns,

additional cost, and potentially introducing radiation concerns for those in the operating theatre [69, 173, 14].

Without the use of intraoperative volumetric scans, the automation of the initial registration between preoperative and intraoperative scans often resorts to the use of artificial cross-modality fiducials to provide additional landmarks that can be automatically localized in all scans [7, 187, 172, 203]. The use of such fiducials creates additional invasive steps in the workflow of typical procedures to the degree that it is often desirable to avoid their use altogether [7, 14]. Furthermore, the utility of the fiducials often diminishes greatly as their placement becomes separated from the surgical surface of interest where their use may be more practical [14].

Lastly, for the update of the registration, the vast majority of techniques rely on some variant of *ICP* for incremental shape-based registration [113]. When combined with the use of typical surface error metrics, such as *SRE*, *ICP* will directly minimize the error metric. However, when combining the relatively smooth surfaces of typical surgical sites with the typical issues of low surface coverage with intraoperatively acquired 3D surface data, a large degree of local shape similarity is present in the problem [173, 158]. The use of *ICP* will often cause the intraoperatively acquired surface to "slide" along the preoperative surface, potentially away from a physically meaningful initial alignment, to minimize the error metric.

Ultimately, many registration techniques have been applied to this problem area; however, no clear cut winner has emerged when considering both technical and practical challenges in the deployment of a final solution. As a result, manual registration between the preoperative and intraoperative datasets is still the de-facto standard in current implementations. Techniques often differ in how they generate this initial registration and how they generate the information necessary

to maintain the registration in the presence of incoming intraoperative data.

2.2.2 Non-rigid deformation

Another area of this field that has room to grow is the ability to model non-rigid deformations in the modelling and registration process. While this issue has begun to be explicitly dealt with in the *SN* literature, many systems still rely on a rigid assumption during one or more stages of the pipeline [113]. For stereo based systems where instantaneous depth measurements are available, incorporation of non-rigid deformation models will likely occur during a registration process (either temporal registration of the reconstructed data or registration of intraoperative information back to preoperative models). Biomechanical models have been explored in the medical imaging community as a potential solution to this problem (e.g., [2, 59, 22, 6, 125]). Due to the large degree of intra- and inter-patient variability in tissue properties, the use of biomechanical models is typically a highly patient specific process [14]. Acquiring meaningful measurements of patient-specific tissue properties in a manner that is suitable for surgical workflow remains a challenge [14]. Such models are also expensive to compute from a computational point of view often making them not real-time amenable in their current state [14].

For motion-based structure recovery techniques, accounting for non-rigidity is likely to occur throughout the entire system. Given that motion-based systems predominantly rely on tracked features as the basis for the estimation of 3D structure and motion, consideration of the features and how they are matched may be part of the process for dealing with the non-rigidity of the scene. Currently, the types of features and descriptors that are present in the medical image-based structure

recovery literature tend to be, at most, invariant to changes due to scale and rotation, but lack invariance to the affine transformations present during non-rigid deformations. As the ability to recover stable features is currently an issue for such systems, consideration of a more stable feature detection/tracking method that can cope with affine transformations may prove useful (e.g., [120, 23, 164]). Further integration of non-rigid techniques present in the computer vision *SfM* literature may also yield benefits for medical applications, where the effects of non-rigidity cannot be avoided, minimized or compensated for via current methods.

2.2.3 Temporal consistency

In the medical stereo literature, postprocessing was found to play a significant role in providing finished results. Often, postprocessing steps were put in place to deal with issues of spatial and temporal smoothness [151, 152, 153, 11]. Varying degrees of enforcing spatial smoothness in an integrated manner during the depth estimation process is already present in the medical stereo literature [193, 182, 146, 151, 11]. Approaching temporal smoothness in an integrated fashion using techniques already present in the stereomotion literature, e.g., [177], would then be an intuitive next step for the field. As many downstream applications for depth estimation depend on the temporal consistency of the depth estimates, such an approach may greatly benefit the abilities of such systems.

2.2.4 Evaluation

Perhaps the largest open issue in this field is the lack of standardization in the evaluation of proposed systems. Currently, proposed systems predominantly use

independently collected information for evaluation. The data used for testing can originate from a variety of sources ranging from simulations to real data of phantoms or biological tissues imaged during in vivo or ex vivo interventions. Evaluation is also often performed seemingly based on whatever data is available given the particular application area, procedure protocol, access available to the researchers, etc. This state of evaluation leads to a vast range of data sources and metrics being used for the evaluation of vision-based surface reconstruction and medical *SN* systems. The vast range of data sources present in the community is, in large part, due to privacy issues that often impedes the sharing/dissemination of real surgical data. Thus, proposed systems are examined predominantly in individual silos with very few conclusions being drawn regarding comparison between systems.

Many sub-fields in the computer vision literature have received a large amount of traction, and experienced sustainable growth, by providing standardized datasets along with standard metrics for evaluation (relevant examples include: two-frame stereo [165], visual odometry [52], object recognition [41], and action recognition [84]). These datasets provide a common ground for comparison between systems. Making a dataset that is suitable for the entire field of *SN* research is a difficult endeavour. Many aspects should be considered, primarily the drastically different tissue properties being imaged during different procedures and across patients. Additionally, many other aspects vary across systems and surgical applications including the varying imaging modalities and the additional constraints being used by different systems to help make individual system approaches more stable. Even considering these challenges, useful datasets should be possible that can focus on a subset of the field that will encourage growth in the knowledge of the constituent components that will be common across all such systems. Furthermore,

consensus on what process and what metrics should be used to evaluate a system would go a long way towards standardization in evaluation of *SN* systems, even for cases where standard data does not yet make sense.

Another issue that could be addressed in this realm would be the inclusion of additional standardized error metrics. Typical measures of performance (e.g., disparity error, *Surface Registration Error (SRE)* and *Target Registration Error (TRE)*) capture aspects of the accuracy of the surface reconstruction and surface registration processes. For the intended applications of this technology, the structures of interest are often suspended in the volume. Metrics that relate accuracy of measurements at the surface to areas of interest in the organ volume will likely be a necessary step for the eventual adoption of this technology. Along similar lines, it is highly desirable to develop specific accuracy and precision requirements for recovered 3D information to be applicable to specific surgical tasks. Ultimately, performance in this domain must relate to the medical needs.

2.3 Comparison of the proposed system to previous work

In light of the previous work, the main contribution of this dissertation is in the design and instantiation of a novel end-to-end *SN* system for application to open liver surgery. Existing *SN* systems that target an open setting tend to rely on intraoperative sensing modalities that require a significant amount of time to generate an intraoperative 3D surface estimate [26, 32, 87, 85] resulting in a "snapshot" approach to the *SN* system as a whole. As such, a pressing need

for *SN* systems targeting open liver surgeries is to be able to provide real-time localization feedback that updates on the current state of the patient [175]. The system presented in this dissertation follows **Path 3** (surface registration of a partial intraoperative 3D surface reconstruction and the preoperative volume) of the functional pathways classification presented in Section 2.1 and in Figure 2.1. The application of this design to open liver surgery removes the need for a "snapshot" approach to *SN* along with any requirement to operate at breath hold and allows for frequent updates. Other design choices within this class of systems were made that respond to pressing needs of such *SN* systems applied to open liver resection tasks.

The presented system uses a hand delineated mask to isolate the liver. Masking techniques have shown up in the extant systems review primarily to guide monocular surface reconstruction efforts (e.g., [37]). Masking in this system allows for the successful application of rigid registration techniques on a scene that is highly non-rigid as a whole for the interframe motion estimation techniques.

The system integrates an adaptive *Coarse-to-Fine (CTF)* stereo algorithm [179]. The use of a *CTF* strategy provides efficiency and added robustness to local minima. *CTF* stereo techniques have been applied favorably to imagery of liver phantoms in an isolated manner [189]. This work applies a similar strategy to both controlled laboratory datasets as well as intraoperative cases, within an end-to-end *SN* pipeline, with the added application of adaptive windowing techniques to help avoid the smoothing across 3D boundaries that is typical of *CTF* approaches [50, 179].

Manual initialization strategies are a common bottleneck in the presented *SN* systems and have been shown to interrupt surgical workflow and have outcomes that are highly dependent on user expertise [14]. Semi-automatic approaches have

appeared in the literature that require the user to hand select unique 2D and/or 3D landmarks that are identifiable in both the intraoperative and preoperative imagery [38, 139, 149]. The proposed semi-automated initialization scheme removes direct manipulation of the data and turns the initialization into a selection process whereby the n-best registration candidates are presented to the user for selection of the most accurate placement of the two surface estimates, further attempting to reduce the manual effort and expertise required for initialization.

Regarding motion estimation, a recent survey noted that, to their knowledge, all shape-based intraoperative registration methods, used in laparoscopic interventions, relied on *ICP* or one of its variants [113]. Systems in the class that perform pose estimation from 2D to 3D point correspondences often make use of landmarks that were selected manually [129, 65] or automatically [143] along with subsequent tracking in order to maintain the estimated projection with incoming video frames. The system presented in this dissertation applies a robust feature-based approach to interframe 3D motion estimation in order to encourage the registration to bring physically meaningful features into alignment and does so without chaining multiple incremental registrations that can rapidly lead to drift (typical of the *ICP*-based approaches prevalent in shape-based approaches to *SN*). The use of a feature-based technique for interframe motion estimation in the *SN* system presented in this dissertation perhaps provides the biggest divergence (not related to the particular surgical application) in comparison to other extant systems belonging to **Path 3** of the functional pathways classification.

2.4 Summary

This chapter has presented a review of existing approaches to image guided *SN* systems in the literature. A system classification was presented categorizing existing systems into four broad categories: (i) systems that perform pose estimation from 2D-3D point correspondences, (ii) systems that capture a model/map intraoperatively and subsequently estimate pose relative to this map, (iii) systems that perform a surface registration of an intraoperatively acquired partial surface reconstruction and the preoperative volume, and (iv) systems that register using an intraoperatively acquired volumetric image as a bridging modality. The most prevalent approaches in the literature are those that perform shape-based registration back to a preoperative volume. Recently, systems that rely on laparoscopically acquired image sequences have increasingly resorted to following an architecture similar to the second category: building a coherent model from multiple frames, registering this model to the preoperative model and then localizing with respect to the intraoperatively acquired surface. It has been noted that the lack of surface coverage in the individual frame reconstructions recovered in these scenarios makes registration directly to the preoperative model a prohibitively difficult task in the majority of cases [173, 158, 215].

Machine Learning (ML) has recently had an impact on the *SN* literature. Examples where learned components for depth estimation [104, 110], organ segmentation [211], and similar view detection [202] have been integrated into *SN* systems has already appeared in the literature. Given the trends in *SN*-related fields towards increased exploration of *ML* techniques, further integration in research-based *SN* systems seems inevitable. However, integration of *ML* components into *SN* systems

may experience further barriers for translation into clinical settings. Outstanding issues include development of technical approaches that allow systems to learn and adapt to specific deployments (e.g., specific hospital settings) and resolving matters related to privacy of patient data used in training and testing [144].

SN applications of computer vision have a history of taxing extant algorithms, representations and systems. While the problem is undoubtedly challenging, great strides have been made to tackle many of the necessary challenges. Advances along the lines outlined under open issues will help not only with *SN* applications, they will undoubtedly contribute to the basic science of the field. This chapter also highlighted how the system that is documented in the remainder of this dissertation contributes to the state-of-the-art.

Chapter 3

Technical approach

This chapter details a novel end-to-end system for fast surface reconstruction and motion estimation for alignment with a preoperative *Computed Tomography (CT)* scan with application to liver surgery. Specifically, this system is deemed to be "end-to-end" as the system implements a fully functional *Surgical Navigation (SN)* pipeline that relies on manual input for initialization purposes only and requires no subsequent user intervention. The current instantiation of the system is designed for use during intraoperative planning/exploratory phases. Targeting this particular phase of surgery provides an aid to help the surgical team relocalize subsurface structures of interest while avoiding the need to deal with the significant occlusions, deformations or changes to the surface topology that typically occur during the course of a typical liver resection task (many of which are still significant challenges in the *Surgical Navigation (SN)* research community). Nevertheless, information is maintained throughout the system, with these future challenges in mind, that could benefit future extensions of this work allowing the system to work in more stages of the surgery.

The system was designed (and validated) for use in an open surgical environment. Despite a shift in surgery towards less invasive procedures [191] and corresponding research effort in *SN* systems that target *Minimally Invasive Surgery (MIS)* applications, as evidenced by the focus of recent surveys in the area [96, 91, 128, 121, 113, 14], most liver resections are performed as open resections due to the extent and location of disease [166].

Underlying the system design is an effective decomposition of the intraoperative field of view to allow for the use of rigid constraints. The surgical field that is in view from the overhead cameras typically includes several surgical surfaces (i.e., the diaphragm, intestines, fatty tissue, outer abdominal cavity, etc.). When looking at the scene in its entirety the movement observed is highly non-rigid. During intraoperative planning stages the majority of the motion of the surface of interest, i.e., the liver, is moving primarily due to the cardiac and respiratory cycles. In this phase the motion of the liver is semi-rigid [173] with typically large translations caused by the respiratory cycle, $\approx$10 - 75 mm [33], constituting the majority of the movement of the liver surface. Similar to previous work targeting the uterus [37], our work proposes the use of a mask that allows for separation of the liver from the surrounding tissues in a single frame. This masking technique allows for rigidity assumptions to be applied on the 3D structure of the liver surface in a way that provides plausible registration results within typical tolerances required for surgical adoption (where errors are often required to be smaller than 5 mm [14]).

As noted in the penultimate section of Chapter 2, Section 2.3, the system includes a number of components that address several needs in the medical community that were not previously addressed or often overlooked in the *SN* research community. Here, key components are threefold. First, the system employs a

novel semi-automated approach to initialize the intraoperative to preoperative data registration process. This component decreases reliance on difficult manual registration initialization. Second, the system integrates an adaptive *Coarse-to-Fine (CTF)* stereo algorithm [179] capable of fast, accurate, and boundary preserving 3D surface reconstruction of the visible anatomy. With important information for downstream processing occurring at 3D boundaries (such as organ contours for pre- to intraoperative registration) this is an important distinction that we believe is a first in the *SN* community. Third, the system employs a 3D motion estimator based on interframe feature matching to register a time series of reconstructions to the initial frame for subsequent registration to a volumetric *CT* scan. This approach encourages registration to bring physically meaningful features into alignment and does so without chaining multiple incremental registrations that can rapidly lead to drift.

This chapter is laid out as follows. Section 3.1 provides an overview of the *SN* system design. Sections 3.2 - 3.5 detail the various sub-components of the *SN* system. Section 3.6 provides implementation details for the instantiated system. Section 3.7 provides a summary of the chapter. The description of the system in this chapter is provided in a way that is independent of the specific parameter values used during deployment. Implementation details for all aspects of the system, and all parameter values used during system evaluation, are provided in Appendix A.

3.1 System overview

Figure 3.1 provides an overview of the system for recovering a time series of 3D surface reconstructions of a surgical scene and subsequent registration of these

reconstructions to a preoperative volumetric model. The input is a stream of synchronized left-right image pairs from a stereo video camera and a volumetric model (a segmented *CT* scan of the patient's liver). The specific imaging system used for evaluation is documented in Section 4.1.

The processing pipeline consists of four main components: system initialization, stereo correspondence determination, 2D feature tracking, and six-*DOF* motion estimation. *System initialization* produces a registration between the initial frame of the intraoperative video and the preoperative volumetric model. *Stereo correspondence* yields a dense disparity map between points in the left and right images. The disparity map is projected into 3D space and filtered to produce a 3D surface reconstruction. *2D feature tracking* is applied to the video from the right camera to provide 2D matched feature locations across the image sequence. These 2D tracks are fused with the disparity maps to produce a 3D non-rigid deformation field. Six-*DOF motion estimation* computes a rigid six-*DOF* motion estimate from the 3D deformation field relating the current frame back to the initial frame. The 3D surface reconstruction and six-*DOF* motion estimate are combined to place the surface into the same reference frame as the initial frame of the video sequence. When combined with an initial registration of the initial frame to the preoperative scan, the registration of incoming frames to the initial frame allows for a chained registration back to the preoperative volumetric scan of the organ of interest.

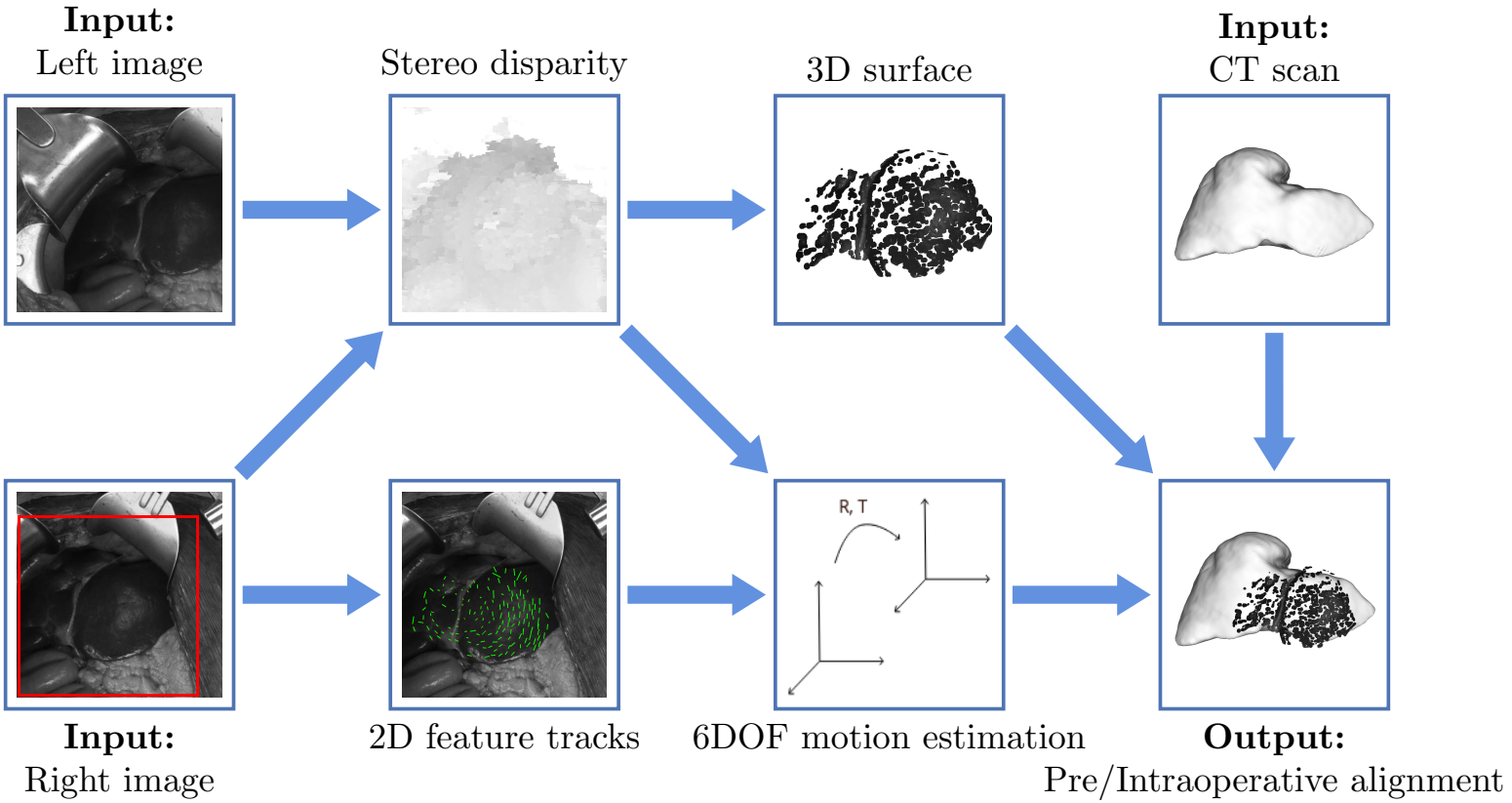


Figure 3.1: System Diagram. Input images from stereo video cameras are shown in grey level with the red box in the right image indicating the region of analysis. Recovered stereo correspondences are shown as a disparity map referenced to the right image, with brighter grey levels indicating larger disparities. Green lines on tracking output show recovered 2D displacement of features across time. Pre/intraoperative alignment shows an overlay of the reconstructed surface map on the corresponding volumetric model. See text for detailed description.

3.2 Initialization

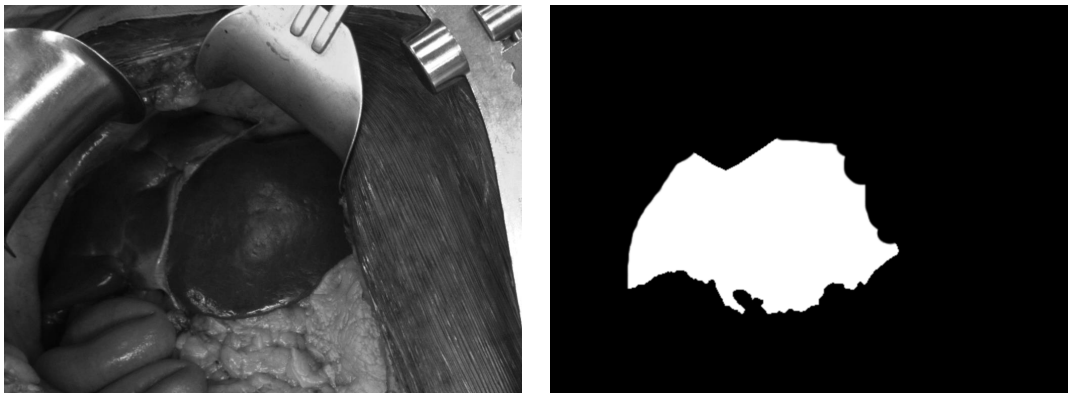
System initialization consists of two areas of user intervention. The first specifies a mask for the image region that covers the organ of interest, i.e., the liver. The second specifies an initial registration between the preoperative CT volumetric image and the intraoperative reconstruction at the initial frame; two approaches are provided for this component.

3.2.1 Region of interest specification

The initialization process starts by isolating the portion of the surgical scene belonging to the liver in the initial frame of the intraoperative video. Scene decomposition is achieved by having a binary mask generated by hand denoting portions of the visible scene that should be reconstructed/tracked by the SN system.

The mask can be created in any suitable image editing software, where the portion of the scene to be recovered is painted in white and the rest of the scene is painted black, see Figure 3.2. For example, the case shown in Figure 3.2 was generated using Photoshop[1]. Use of the mask, $M^{t^*}(x, y)$, where (x, y) denotes the image spatial position and t^* specifies the frame in the video where initialization occurs (the initial frame), allows for the subsequent 3D reconstruction and tracking procedures to accurately decompose the scene and focus on the liver. A hand generated mask is only provided for the reference (right) image and only at the initial frame, t^*.

[1] Adobe Photoshop is a graphics editor developed and published by Adobe Inc., San Jose, CA, United States.



(a) (b)

Figure 3.2: Region Of Interest (ROI) Specification. (a) Shows the reference (right) image in the initial frame of the intraoperative sequence used during initialization. (b) Shows the hand-delineated mask marking the portion of the liver surface that will be reconstructed and tracked by the developed *SN* system.

3.2.2 Manual registration initialization

The first registration initialization strategy is a manual 3D-3D surface registration, where the user takes the 3D surface model of the liver, provided by a segmented *CT* scan, and aligns it with the partial 3D surface model reconstruction at the initial frame through manual use of a *Graphical User Interface (GUI)*, MeshLab [31], see Figure 3.3. The hand alignment is refined by the use of *Iterative Closest Point (ICP)* to minimize any local error in the initial hand alignment and is reinspected to ensure the alignment is still physically meaningful [15, 29, 212]. This initialization strategy is in-line with common practice in the *SN* community and represents a comparative measure to most existing systems for these tasks. Performing this task requires both knowledge of the anatomy and comfort with the imaging modalities and experience with *Computer-Aided Design (CAD)/Computer-Aided Manufacturing (CAM)* type interfaces for working with 3D data. As such, manual registration

typically takes some time and effort to perform with accuracy, even for an expert operator.

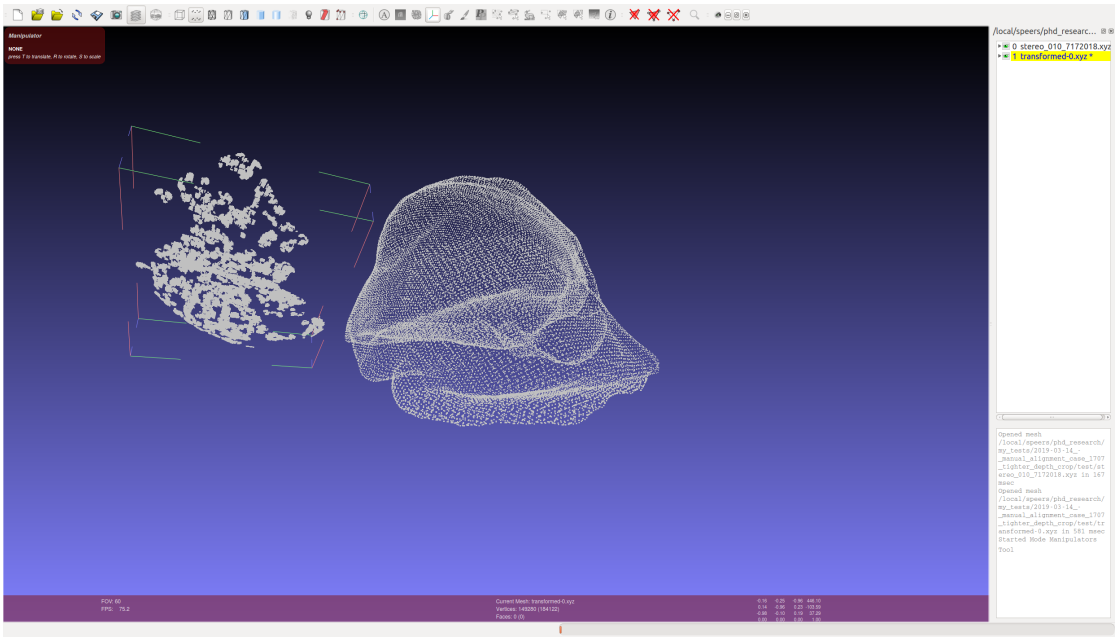


Figure 3.3: Manual Registration Initialization. The figure shows an intermediate view of using MeshLab during the manual registration process. Points to the left side correspond to the intraoperatively acquired surface model (point cloud). The mesh on the right side corresponds to the preoperatively acquired volumetric model.

3.2.3 Semi-automated initialization

The second initialization strategy automates all or part of the initial alignment relating the intraoperative 3D reconstruction with the preoperatively acquired surface estimate of the patient's liver. *ICP* [29, 15] is perhaps the most well-known and widely used method for efficiently registering 3D point clouds under Euclidean (rigid) registration, both in terms of initialization strategies in *SN* systems as well as in broader usage scenarios requiring registration of 3D point sets. The algorithm

starts with an initial transformation estimate and then begins an iterative procedure alternating between finding closest point correspondences between the two point clouds under the current transformation and re-estimating the transformation using these closest point correspondences, iterating until convergence. The primary downside to this approach is its susceptibility to converging to a local minimum due to the non-convexity of the problem and the iterative nature of the *ICP* approach. Thus, *ICP* requires a reasonably good initialization to have the best chance at converging to a physically meaningful location. The relatively flat solution space caused by the registration of smooth surfaces typical in *SN* applications creates additional issues for *ICP* when attempting to converge to an exact position on shape-based information alone.

To help decrease reliance on needing to provide a manual initialization, several global registration methods have been developed in the broader literature. One approach to this problem is to perform a stochastic optimization such as genetic algorithms [170, 148], particle swarm optimization [196], particle filtering [161] and simulated annealing schemes [19, 135]. While these approaches help to avoid local minima, initializations still need to be reasonably good otherwise the parameter space can become too large for typical heuristic search. These solutions are also incapable of ensuring they reach a globally optimal solution.

Another approach to the problem uses shape descriptors that can be extracted and matched between the two point sets allowing for a coarse alignment of the two point sets. Local descriptors such as spin images [82], shape contexts [10], integral volume [53] and *Fast Point Feature Histograms (FPFHs)* [159] are invariant to specific transformations and have been used for such purposes. Feature correspondences, and corresponding transformation estimation, can then be determined using

random sampling methods [159], greedy algorithms [82], Hough transforms [205] or *Branch-and-Bound (BnB)* algorithms [53, 9]. Global shape descriptors, such as *Extended Gaussian Images (EGIs)* [114], have also been used to find an optimal transformation maximizing global descriptor correlation but require significant surface overlap to perform reasonably. The main drawback of these approaches is that they require that distinctive 3D surface texture be present on the surface of the organ of interest that can be faithfully recovered in both the preoperative model and the intraoperatively acquired surface reconstruction to guide the registration. This prerequisite was not met in any of the laboratory or intraoperative evaluations performed in this work. As a result, such local feature-based approaches were not considered for the semi-automated initialization technique presented in this dissertation.

The semi-automated initialization strategy used in this pipeline employs a modified version of *Globally Optimal Iterative Closest Point (GO-ICP)* [208] to provide registration candidates to a user. A fixed number of registration candidates are computed and displayed to the user allowing them to pick the best option from the set, as shown in Figure 3.4. The task then becomes one requiring knowledge of the anatomy and familiarity with viewing this type of data, skills typically abundant in the surgical environment. Perhaps more importantly, this strategy removes the burden of having to directly manipulate the 3D data in a meaningful way, a skill that would be at a premium in both the surgical environment and workflow.

Traditional *ICP* takes two point sets, $\mathcal{Y} = \{\mathbf{Y}_i\}, i = 1, \ldots, G$ and $\mathcal{Z} = \{\mathbf{Z}_j\}, j = 1, \ldots, H$ and attempts to find the rigid rotation, R, and translation, $\mathbf{T}$, that

minimize the L_2-error E_{ICP},

$$E_{\text{ICP}}\left(\mathsf{R},\mathbf{T}\right)=\sum_{i=1}^{G}e_i\left(\mathsf{R},\mathbf{T}\right)^2=\sum_{i=1}^{G}\|\mathsf{R}\mathbf{Y}_i+\mathbf{T}-\mathbf{Z}_{j^*}\|^2, \tag{3.1}$$

where $e_i(\mathsf{R},\mathbf{T})$ is the per-point residual error for $\mathbf{Y}_i$ as compared to the closest point on the model surface $\mathbf{Z}_{j^*}$. Given R and $\mathbf{T}$, the point $\mathbf{Z}_{j^*}$ is given as,

$$j^*=\underset{j\in\{1,\cdots,H\}}{\arg\min}\|\mathsf{R}\mathbf{Y}_i+\mathbf{T}-\mathbf{Z}_j\|. \tag{3.2}$$

Unlike the standard *ICP*, *Globally Optimal Iterative Closest Point (GO-ICP)* makes use of a *Branch-and-Bound (BnB)* strategy [92] to find a globally optimal shape-based registration under the same L_2-error used in the traditional *ICP* implementations, (3.1). *BnB* strategies are often used in solutions to non-convex problems (such as the 3D point set registration problem). Application of *BnB* techniques to 3D registration adds the challenges of finding a reasonable quantization of the search space as well as defining the upper and lower bounds of areas in the quantized search space. The translation and rotation search spaces are maintained as separate 3D cubes (further quantized into uniform sub-cubes) resulting in a translation and rotation vector (where rotations are computed in angle-axis format). The upper bound in a given translation/rotation sub-cube pair is given as the L_2-error evaluated at the center point of the candidate sub-cubes and the lower bound is determined by the upper bound and the uncertainty radius around this upper bound; at each point the uncertainty radius is a function of the length of a side of the sub-cube in the quantized search space and the distance of the point to the origin. The full six dimensional search space is evaluated using a nested *BnB* strategy whereby an outer *BnB* searches the rotation space of $SO(3)$

and solves the upper and lower error bounds, $\bar{E}_R$ and $\underline{E}_R$ respectively, of a given candidate rotational sub-cube, C_R, and the corresponding optimal translation, $\mathbf{T}_{\mathbf{C_R}}$, by executing an inner *BnB* search. Both the inner and outer *BnB* search processes use a queue to keep track of sub-cubes that need to be processed in the translational, Q_T, and rotational search spaces, Q_R, respectively. Each search space is initialized with a single cube encompassing the entire rotational or translational search space to be considered. At each iteration, this search space is split into eight equal sub-cubes that are individually run through an update procedure. The goal of the search is to find a globally optimal L_2-error, E^*, and corresponding optimal rotation, R^*, and translation, $\mathbf{T}^*$. Given a candidate rotational sub-cube, C_R, the update procedure is shown in Algorithm 1.

The default *GO-ICP* implementation works well for scenarios where the surface characteristics of the model combined with the surface coverage observed in the data to be registered to the model allows for a unique match [208]. Unfortunately, given the rather low surface coverage typical in intraoperative stereo reconstructions (often capturing only a portion of one face of the liver), the large degree of local shape similarity present in the model and potential for deformation between preoperative an intraoperative states, the single candidate provided by *GO-ICP* is typically unlikely to be of use in intraoperative settings. As such, the algorithm was modified to maintain a list of candidate registrations. By modifying the *BnB* conditions in the algorithm, the n-best registration candidates can be produced using the same search strategy. If the number of registration candidates to be considered is known a priori then the list of registration candidates can be maintained during the search process. The modified update procedure in this case is shown in Algorithm 2. In the use case where the operator would like to generate an additional registration candidate

the update procedure can be modified to provide the "next best" registration candidate at the cost of some overlapping computation. The update procedure in this case is shown in Algorithm 3. During initialization, either Algorithm 2 can be used to generate the n-best solutions allowing for the selection of a suitable registration from a set of candidate registrations or the user can request candidates one at a time getting the "next best" registration candidate using Algorithm 3 until an acceptable registration is provided. For evaluation purposes, the 50-best registration candidates were generated (using the approach outlined in Algorithm 3) and inspected for each of the datasets evaluated in Chapter 4 as preliminary experimentation suggested that the solution would be within that group.

Algorithm 1: Globally Optimal Iterative Closest Point (GO-ICP) update procedure for candidate rotational sub-cube

Input: Candidate rotational sub-cube C_R, current minimal error E^* (and corresponding rotation R^* and translation $\mathbf{T}^*$) and priority queue of candidate sub-cubes Q_R.

Output: Updated minimal error E^* (and corresponding rotation R^* and translation $\mathbf{T}^*$) and priority queue of candidate sub-cubes Q_R.

Compute upper-bound error $\bar{E}_R$ for candidate rotational sub-cube C_R and corresponding optimal translation $\mathbf{T}_{\mathbf{C_R}}$ by calling an inner translational *BnB* with rotation R_0, zero rotational uncertainty radii, and a current minimum error E^* (where R_0 is the central point of the candidate rotational sub-cube).

if $\bar{E}_R < E^*$ **then**

    Run *ICP* with the initialization $(R_0, \mathbf{T}_{\mathbf{C_R}})$. Update E^*, R^*, and $\mathbf{T}^*$ with the results of *ICP*.

Compute $\underline{\mathrm{E}}_R$ for C_R by calling an inner translational *BnB* with rotation R_0, uncertainty radius γ_R and current minimum error E^* (where γ_R is the uncertainty radius for rotational sub-cube C_R, i.e., the maximum amount a point $\mathbf{x}$ can be perturbed by an arbitrary rotation $R \in C_R$).

if $\underline{E}_R \leq E^*$ **then**

    Discard C_R and continue evaluating sub-cubes in the priority queue Q_R.

else

    Put C_R into Q_R.

Algorithm 2: Globally Optimal Iterative Closest Point (GO-ICP) "list-based" update procedure for candidate rotational sub-cube

Input: Candidate rotational sub-cube C_R, **an ordered list of size** n **containing the best errors evaluated at this point in the search process** E^* **(and corresponding rotation** R^* **and translation** $\mathbf{T}^*$ **lists with matching indices)** and priority queue of candidate sub-cubes Q_R.

Output: **Updated error list** E^* **(and corresponding rotation** R^* **and translation** $\mathbf{T}^*$ **lists)** and priority queue of candidate sub-cubes Q_R.

Compute upper-bound error $\bar{E}_R$ for candidate rotational sub-cube C_R and corresponding optimal translation $\mathbf{T}_{\mathbf{C_R}}$ by calling an inner translational BnB with rotation R_0, zero rotational uncertainty radii, and a current minimum **target error** $\max(E^*)$ (where R_0 is the central point of the candidate rotational sub-cube).

if $\bar{E}_R < \max(\mathbf{E}^*)$ **then**

    Run ICP with the initialization $(R_0, \mathbf{T}_{\mathbf{C_R}})$. Update E^*, R^*, and $\mathbf{T}^*$ with the results of ICP **by inserting the error (and corresponding rotation and translation values) into the corresponding lists at matched indices sorted by the error measurement.**

Compute $\underline{\mathrm{E}}_R$ for C_R by calling an inner translational BnB with rotation R_0, uncertainty radius γ_R and current minimum **target error** $\max(E^*)$ (where γ_R is the uncertainty radius for rotational sub-cube C_R, i.e., the maximum amount a point $\mathbf{x}$ can be perturbed by an arbitrary rotation $R \in C_R$).

if $\underline{E}_R \leq \max(\mathbf{E}^*)$ **then**

    Discard C_R and continue evaluating sub-cubes in the priority queue Q_R.

else

    Put C_R into Q_R.

Algorithm 3: Globally Optimal Iterative Closest Point (GO-ICP) "next best" update procedure for candidate rotational sub-cube

Input: **Error from previously obtained registration candidate** E^+, candidate rotational sub-cube C_R, current minimal error E^* (and corresponding rotation R^* and translation $\mathbf{T}^*$) and priority queue of candidate sub-cubes Q_R.

Output: Updated minimal error E^* (and corresponding rotation R^* and translation $\mathbf{T}^*$) and priority queue of candidate sub-cubes Q_R.

Compute upper-bound error $\bar{E}_R$ for candidate rotational sub-cube C_R and corresponding optimal translation $\mathbf{T}_{\mathbf{C_R}}$ by calling an inner translational *BnB* with rotation R_0, zero rotational uncertainty radii, and a current minimum error E^* (where R_0 is the central point of the candidate rotational sub-cube).

if $\bar{E}_R < E^*$ ***and*** $\bar{E}_R > E^+$ **then**

    Run *ICP* with the initialization $(R_0, \mathbf{T}_{\mathbf{C_R}})$.

    if ***Resulting error from*** **ICP** $> E^+$ **then**

        Update E^***,** R^***, and** $\mathbf{T}^*$ **with the results of** ***ICP*****.**

Compute $\underline{\mathrm{E}}_R$ for C_R by calling an inner translational *BnB* with rotation R_0, uncertainty radius γ_R and current minimum error E^* (where γ_R is the uncertainty radius for rotational sub-cube C_R, i.e., the maximum amount a point $\mathbf{x}$ can be perturbed by an arbitrary rotation $R \in C_R$).

if $\underline{E}_R \leq E^*$ **then**

    Discard C_R and continue evaluating sub-cubes in the priority queue Q_R.

else

    Put C_R into Q_R.

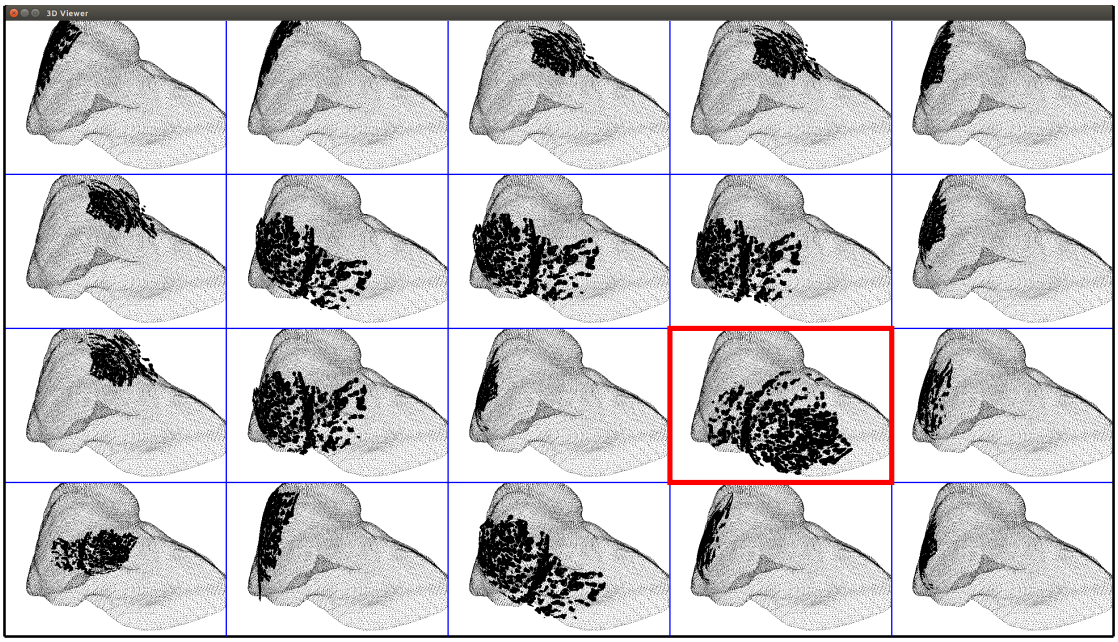


Figure 3.4: Globally Optimal Iterative Closest Point (GO-ICP) Candidate Selection Interface. The top 20 registration candidates are displayed on a 4×5 grid. The best registration candidate is selected by the user and is highlighted in red. The user is able to manipulate the view of all registration candidates in a synchronized manner (i.e., the *CT* scans common to all registration options will all have the same orientation to the camera when the view is manipulated).

3.3 Stereo correspondence and 3D surface reconstruction

3.3.1 Stereo correspondence

Several intraoperative imaging modalities have been used for the purposes of *SN*. Due to the trend towards *MIS*, and prevalence of monocular endoscopic cameras in many surgical disciplines, a large body of work exists in attempting to use monocular intraoperative image sequences in *SN* systems (e.g., [207, 197, 74, 130, 115, 36, 116, 123, 37, 56, 90, 103, 216]). *SN* systems leveraging monocular imagery as their

intraoperative imaging modality have been previously mentioned in Section 2.1 (with particular focus in Sections 2.1.1 and 2.1.2). One method for using monocular imagery in *SN* applications is to maintain a projection relating the preoperative volumetric model to the current image (e.g., [7, 187, 172, 123, 129, 142, 65, 77]). This projection can be estimated with a set of 2D-3D point correspondences being generated and/or maintained between the two image sources. One major hurdle in these applications is the relative paucity of features that can be uniquely and precisely identified between a typical preoperative scan of the patient's anatomy and a 2D image from a monocular video sequence obtained during that patient's surgery. Another approach is to perform pose estimation with respect to a model or map captured intraoperatively. These methods require the generation of an intraoperative 3D surface estimate of the patient's anatomy from the monocular image sequence. *Structure-from-Motion (SfM)* (e.g., [207, 197, 74, 115, 123, 90]), *Simultaneous Localization and Mapping (SLAM)* (e.g., [130, 37, 56]), *Shape-from-Shading (SfS)* (e.g., [36, 116]), *SfS+SfM* hybrids (e.g., [116]) as well as learning-based approaches (e.g., [103, 216]) have all appeared in the *SN* literature as methods for acquiring intraoperative 3D surface estimates from monocular image sequences. The vast majority of these techniques require strong assumptions to be met during the acquisition of the models (examples include assuming scene rigidity, specific camera motion, and/or controlled lighting combined with the typical use of an oversimplified surface reflectance model for the given domain). These assumptions are, in general, a result of the fact that depth estimation from a single view is an ill-posed problem. Specifically, the geometry of the problem of recovering 3D structure from a monocular image only allows for each viewed point to be constrained to a line in space without making use of additional information

and/or assumptions about the scene geometry being viewed. Thus, while depth information can not be generated directly from each independent frame in the sequence, after the intraoperative scene geometry has been recovered, incoming frames observing the scene have the potential to be localized to an intraoperatively obtained map and/or the preoperative model (with the appropriate initialization) at frame rate.

Structured light systems have been utilized to some extent in the *SN* literature in an attempt to further extend the range sensing abilities of traditional image sources through the use of active sensing techniques. Structured light systems project known patterns onto the scene and recover that projected pattern from one or more cameras observing the scene. The projector, in effect, acts as an additional camera in the system, with disparity being calculated between the undistorted pattern displayed by the projector and the distorted projection captured by the camera(s). Structured lighting systems are able to produce relatively dense depth estimates at frame rate (e.g., VGA resolution depth images at 30 Hz for the structured light based Microsoft Kinect[2]). Unfortunately, structured light techniques often suffer in situations where there are other strong light sources present, highly reflective surfaces and/or (semi-)transparent surfaces being imaged [162], all of which are commonplace in typical *SN* applications. Furthermore, development of custom structured light devices intended for *SN* applications also have the additional drawback of typically operating in the visible portion of the spectrum, thus projecting visible patterns onto the patient's anatomy causing an undesirable distraction to the surgical team the device is intended to aid. Potential for encoding

[2]The Kinect is a motion sensing input device for the Xbox video game console produced by Microsoft Corporation, Redmond, WA, United States.

structured light patterns in a manner that is imperceptible to a human operator, and is independent of the carrier signal, has been explored [35]. This approach could be beneficial for future visible-light structured light systems targeting *SN* potentially allowing for the projector to act as a fully adjustable light source for the surgeon while additionally enabling real-time dense active range sensing.

Another intraoperative imaging modality that has been used in *SN* applications is that of *Laser Range Scanning (LRS)* [33, 175, 55]. *LRS* is an active imaging technique that works by sweeping a single-plane laser across the surgical scene. The projection of this laser stripe is captured by a high resolution camera. Through the process of triangulation, the data acquired by the camera can be used to generate 3D point locations along the projected line resulting in the computation of a 3D point cloud of the observed scene during the sweeping of the laser across the scene. The process of sweeping the laser across the scene results in the 3D reconstruction process using *LRS* to be a time aggregated process (taking roughly 30 s for a single acquisition of the liver [175]). As such, systems relying on *LRS* as a 3D reconstruction technique are only applicable for a "snapshot" approach to *SN* where results are shown at a single point in time as opposed to a system that updates with the changing state of the patient's anatomy. Moreover, changes in the surface geometry during a scan (e.g., due to breathing) can yield nonveridical reconstructions.

Based on the above considerations, the system outlined in this dissertation uses stereo video as the intraoperative imaging modality used to register the preoperative and intraoperative states of the patient's liver. Stereo video has a long history of being used in *SN* applications (e.g., [42, 193, 185, 151, 152, 153, 60, 188]). In general, the use of stereo imagery helps to alleviate some of the issues when recovering 3D

structure from monocular imagery by providing multiple synchronized views of the same scene from different points of view. With known calibration, computation of the depth of a point comprises of finding matching locations in the two stereo images and subsequent triangulation. Consequently, dense disparity estimates can be directly computed at each frame in a passive manner placing less constraints on the scene, lighting and movement of the camera.

Existing systems have sampled a wide variety of stereo algorithms for use on various disparate medical solutions. Stereo correspondence techniques primarily differ based on the degree and nature of the spatial smoothness constraints applied to the resulting disparity estimates. Local matching approaches have been explored in the medical imaging literature (e.g., [183, 12]). These approaches are able to provide dense depth estimates over the reference image frame and are typically computationally efficient and real-time amenable due to their restriction to local computation when determining correspondences between the stereo imagery. These solutions are often paired with postprocessing techniques to provide spatial smoothness in the depth estimates. Alternatively, spatial smoothness constraints can be integrated into the correspondence determination itself resulting in what are referred to as semi-global and global methods for determining stereo correspondence. Global methods attempt to minimize an energy function over the entire reference image that typically combines a local match component as well as a global smoothness term (e.g., [94, 182, 28, 11]). Optimizing such an energy function gives solutions that are consistent with the image data and also yield smooth surfaces, but at the expense of high computational complexity. Semi-global methods improve computational efficiency by providing an estimate for a global energy function by performing a series of local estimates (e.g., [193, 151, 152, 153]). Given

consideration for the potential future integration of the *SN* system into surgical workflow only local correspondence techniques were considered for integration into the system due to runtime constraints.

The system uses a local stereo correspondence algorithm that has been shown to provide accurate and efficient depth estimates [179]. This particular algorithm has shown to have a good speed vs. accuracy tradeoff in comparison to many other local and semi-global correspondence methods from the broader stereo vision literature [179]. Given a calibrated stereo pair of images (Figure 3.5), $I_l(x, y)$ and $I_r(x, y)$, with subscripts l and r denoting left and right images, respectively, while (x, y) denotes image position, the algorithm yields a disparity map, $d(x, y)$, that provides the spatial offset between corresponding points in the input pair. The disparity map is recovered by solving the optimization problem:

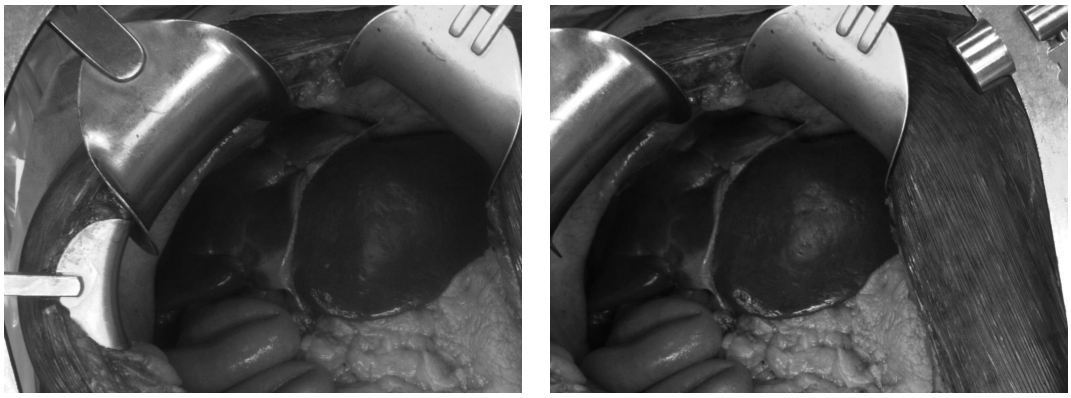
$$d(x, y) = \underset{d_i \in d_{\max}}{\arg\max} \sum_{(u,v) \in w_S(x,y)} \rho \left[I_l(u, v),\ I_r(u + d_i, v) \right], \tag{3.3}$$

via search over disparities, d_i, in a specified range, $d_{\max}$, to maximize the summed pixel-wise similarity measure, ρ, between image intensity values in a window, w_S, around (x, y). Although the algorithm is agnostic with respect to the choice of similarity measure, ρ, our instantiation makes use of *Normalized Cross Correlation (NCC)*,

$$d_{\text{NCC}}(x, y) = \underset{d_i \in d_{\max}}{\arg\max} \sum_{(u,v) \in w_S(x,y)} \frac{I_l(u, v) I_r(u + d_i, v)}{\sqrt{\sum_{(u,v) \in w_S(x,y)} I_l(u, v)^2 \sum_{(u,v) \in w_S(x,y)} I_r(u + d_i, v)^2}}, \tag{3.4}$$

as it is robust to baseline shifts in the image intensity between the left and right images that often occurs in the intraoperative setting.

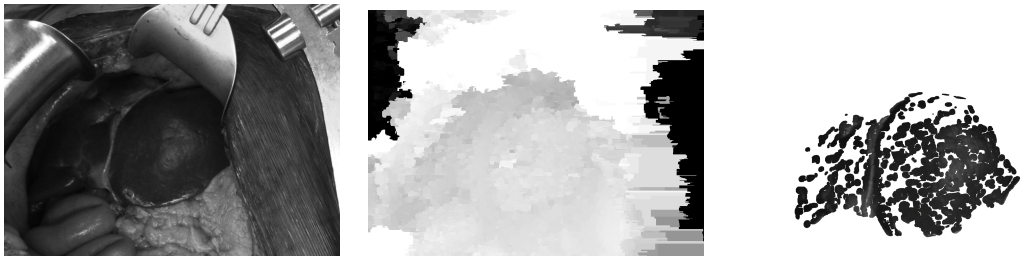
The algorithm employs *CTF* processing for efficiency and added robustness to local minima in the correspondence search, whereby initial low-resolution versions of the input images yield low-resolution disparity maps that subsequently are refined via consideration of higher-resolution images to culminate with the resolution of the original input. The algorithm also uses adaptive windows, w_S, that conform to avoid smoothing across 3D boundaries [50]. An example recovered disparity map for an input stereo pair is shown in Figure 3.6b.



Left Image
(a)

Right Image
(b)

Figure 3.5: Intraoperative Stereo Imagery. Sample imagery taken in an open surgical setting of a liver resection are shown.



Right (Reference) Image
(a)

Disparity Image
(b)

Smoothed 3D Point Cloud
(c)

Figure 3.6: Intraoperative 3D Reconstruction. A reference 2D image of the surgical site is shown in (a). The result of the stereo matching approach is shown in (b). The corresponding point cloud, after all post processing, is shown in (c).

3.3.2 3D reconstruction

Appeal to known camera calibration allows for the recovered disparities to be projected to a 3D point cloud. Given image location $\mathbf{x} = (x, y)$ in pixels, a 4×4 perspective transformation matrix, Q, computed during stereo calibration and the corresponding disparity at the image location, $d(\mathbf{x})$, the 3D reprojection of the point, $\mathbf{\Pi}$, is given as

$$\mathbf{\Pi}(\mathbf{x}, d(\mathbf{x})) = (X, Y, Z)^{\top} = \left(\frac{\tilde{X}}{W}, \frac{\tilde{Y}}{W}, \frac{\tilde{Z}}{W}\right)^{\top}, \tag{3.5}$$

where

$$\begin{bmatrix} \tilde{X} \\ \tilde{Y} \\ \tilde{Z} \\ W \end{bmatrix} = \mathsf{Q} \begin{bmatrix} x \\ y \\ d(x, y) \\ 1 \end{bmatrix} \tag{3.6}$$

and

$$\mathsf{Q} = \begin{bmatrix} 1 & 0 & 0 & -c_x \\ 0 & 1 & 0 & -c_y \\ 0 & 0 & 0 & f \\ 0 & 0 & -1/b & (c_x - c'_x)/b \end{bmatrix}. \tag{3.7}$$

In the above, f is the focal length of the reference camera, b is the baseline distance between the left and right camera, (c_x, c_y) is the principal point in the reference image, and $\left(c'_x, c'_y\right)$ is the principal point in the non-reference image.

The resulting point cloud is filtered via statistical outlier removal to reject

depth values that differ significantly from their neighbours [160]. The sparse outlier removal looks at the k-nearest-neighbours of every point and computes the mean, μ, and standard deviation, σ, of the nearest neighbour distances. Any points falling outside the threshold $\mu \pm \alpha \cdot \sigma$ are identified as outliers and removed from further processing, with α being a scaling factor that was set empirically.

After outlier removal is performed, the remaining points are resampled via *Moving Least Squares (MLS)* [1] to produce a smooth manifold surface as the final 3D surface reconstruction. The goal of this procedure is to define a projection, $P(\cdot)$, that can take a 3D point, $\mathbf{P}_i \in \mathbb{R}^3$, sampled from a surface, S, and project it onto a two-dimensional surface, S_P, that approximates the sampled points, $\mathbf{P}_i$. Given that a manifold surface can be thought of locally as a plane, a local reference domain (plane) is defined for a given point, $\mathbf{R}$.

Notably, statistical outlier removal and *MLS* resampling is only performed for visualization and *Surface Registration Error (SRE)* evaluation purposes. Dense raw 3D reprojection information is used for all motion estimation processes.

3.3.3 Region of interest enforcement

As noted in Section 3.2.1, during initialization a region of interest mask, $M^{t^*}(x, y)$, was provided for the initial (key) frame, t^*, in the image sequence. That mask is used to restrict processing to the area to be registered to the volumetric model. Following initial delineation, the mask is warped automatically to all other frames in the sequence via a robust affine estimate of image motion across the masked region [18]. (Sequences of up to 1729 frames in duration are evaluated in Chapter 4.)

This approach uses an affine model for the parametric flow function,

$$\mathbf{v}\left(x, y; \mathbf{a}\right) = \begin{bmatrix} v_x(x,y) \\ v_y(x,y) \end{bmatrix} = \begin{bmatrix} a_0 + a_1 x + a_2 y \\ a_3 + a_4 x + a_5 y \end{bmatrix}, \tag{3.8}$$

where (x, y) specifies image location, $(v_x(x,y), v_y(x,y))$ are the x and y components of the optical flow at this point which is written in terms of the affine parameters $\mathbf{a} = (a_0, a_1, \ldots, a_5)$. The use of an affine motion model allows for larger windows in the spatial aggregation process used to constrain the computation of the optical flow. When using larger windows, the use of least squares estimation techniques become less appropriate as multiple motions become more likely to be present in a given window. As such, the use of a robust error norm, the Geman-McClure norm [54], λ, results in solving for the affine flow parameters, $\mathbf{a}$, by minimizing the error function,

$$E_{\text{FLOW}}(\mathbf{a}) = \sum_{(x,y)\in w_F} \lambda\left[(\bigtriangledown^\top I)\mathbf{v}(\mathbf{a}) + I_t, \gamma\right], \tag{3.9}$$

where w_F is the support window, $\bigtriangledown = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right)^\top$ is the local gradient vector, $I_t = \frac{\partial I}{\partial t}$ is the temporal gradient and γ is a control parameter. In essence, the error (3.9) simply respects the optical flow constraint equation [73],

$$\left(\bigtriangledown^\top I\right)\mathbf{v} + I_t = 0, \tag{3.10}$$

but generalizes by allowing $\mathbf{v}$ to take parametric form (3.8). The local minima of (3.9) can then be found using *Successive Over-Relaxation (SOR)* techniques [141, 184, 194]. To begin, the support window, w_F, is set to the rectangular *Region Of*

Interest (ROI) encapsulating the marked pixels belonging to the liver in the image mask, $M^{t^*}(x, y)$, provided during initialization, Section 3.2.1. Subsequently, the estimated affine motion parameters, (3.8), are used to propagate the *ROI* (and hence w_F) across time via image warping. The warped masks are currently only used to restrict processing as well as for visualization and evaluation purposes (allowing for the liver to be separated from the surgical scene).

An example recovered surface reconstruction is shown in Figure 3.6c.

3.4 2D feature tracking and generation of a 3D deformation field

The next step in the pipeline maintains a set of 2D feature correspondences across the image sequence that can be fused with the depth information, provided via the stereo correspondence module (Section 3.3), to generate an estimate of the 3D deformation field relating the scene at a given frame back to its corresponding state in the initial frame of the image sequence. The first part of this task, i.e., maintaining a set of 2D feature correspondences across the image sequence, can traditionally be accomplished in two distinct ways: image motion based feature tracking and feature matching. Both of these methods rely on feature extraction, in one or more of the images. These feature locations indicate positions exhibiting distinctive visual texture and can focus on corner-like structures (for detectors such as Harris corners [61], "good features to track" [168], *Features from Accelerated Segment Test (FAST)* [156], or *Oriented Features from Accelerated Segment Test and Rotated Binary Robust Independent Elementary Features (ORB)* [157]) or

blob-like structures (such as *Scale-Invariant Feature Transform (SIFT)* [106] and *Speeded Up Robust Features (SURF)* [8]). Ideal features in this application should be both repeatable (i.e., detected in all images where the feature location appears) and should be able to be precisely located in subsequent images.

Image motion-based feature tracking methods typically use optical flow to obtain the motion vectors relating the location of feature points in adjacent frames effectively propagating the tracked location of feature points extracted at some previous frame in the image sequence. Typical assumptions of these methods include brightness constancy (i.e., the projection of the same point in the scene looks similar in adjacent frames), small motion (i.e., the feature point does not move very far between frames) and spatial coherence (i.e., feature points move like their neighbours). As such, tracking is often more suitable for *SN* systems operating on a video sequence rather than independently captured images that may be taken from very different view points or at different stages of the surgery. The main challenges in tracking-based approaches are dealing with large inter-frame motions, strong illumination changes over time, occlusion, and changes of the appearance of the areas of interest in the surgical scene. These scenarios can cause issues such as drift (incremental build up in the error of the tracked locations of the feature points) and feature drop-off (where tracking is lost for a given feature). Even given these complications, tracking methods can be highly accurate and can capture dense motion fields [198].

Feature matching methods on the other hand attempt to match extracted features between the two images. Feature matching can be performed on the raw intensities using block-matching techniques similar to those used in stereo processing. More often, feature matching is performed on local feature descriptors extracted

around the feature point location in the image. Common feature descriptors include distribution-based descriptors (using histograms to represent different characteristics of the local visual appearance or shape), spatial-frequency techniques (describing frequency content in the image), and differential descriptors (providing properties of local spatial derivatives of the image). These intermediate representations can often allow the matching of these feature descriptors to provide elements of scale and rotation invariance as well as additional robustness to changes in illumination over time. Thus, these methods are best suited to matching features when the observations are made under noticeably different viewing conditions (i.e., changes in view angle, illumination, tissue state, etc). Feature matching methods often provide a very sparse estimate of the motion field. Additionally, subpixel accuracy of matching approaches are often limited with the majority of detectors providing integer precision localization.

For the given application, optical flow-based feature tracking methods were used to maintain the location of extracted features in the initial frame of the image sequence across the entire sequence. Specifically, "good features to track" were extracted and tracked with local estimates of optical flow [168]. "Good features to track" have a long history of use and are widely seen to be reliable features for tracking tasks. Furthermore, "good features to track" have been recently at the core of state-of-the-art action recognition techniques (prior to the introduction of *Machine Learning (ML)* techniques to the field) [198].

To establish feature tracks between frames in the stereo video sequence, 2D features are tracked in one of the video streams (the right is used) in the operative ROI, $M^t(x, y)$, as propagated in time, as specified in Section 3.3.3. Tracking in 2D makes use of the "good features to track" algorithm [168]. This algorithm

restricts operations to sets of 2D feature points, $\chi(t)$, where the local image gradient structure is sufficient for stable appearance across time. Let $\mathcal{K}$ be the set of valid feature indices and $\mathcal{T}$ be the set of valid frame indices. Let a given feature, with feature index $k \in \mathcal{K}$ at frame $t \in \mathcal{T}$, be expressed as $\mathbf{x}_k^t \in \chi(t)$. Tracking is performed on each extracted feature, $\mathbf{x}_k^t$, across two images, I^t and $I^{t+\delta t}$, taken at times t and $t + \delta t$, respectively by minimizing the dissimilarity measure

$$E_{TRACK}\left(\mathbf{x}_k^t\right) = \sum_{(u,v)\in w_T\left(\mathbf{x}_k^t\right)} [I^{t+\delta t}(u+\delta_x, v+\delta_y) - I^t(u,v)]^2, \tag{3.11}$$

over windows, w_T, centred at the feature points. A gradient-based solution is employed to yield the optimal feature displacement, $\mathbf{d} = (\delta_x, \delta_y)^\top$, for each feature point, $\mathbf{x}_k^t$. While the original formulation accounted for a full affine transformation across time [168], the simpler translational formulation given here has proved to suffice for the cases of current interest. Example tracks are shown in Figure 3.7a.

Analogous to the projection of stereo disparity maps to 3D point clouds (Section 3.3), the extracted features at each frame are combined with the calibration information and disparity estimates at the tracked points to yield a set of tracked 3D features, $\mathcal{X}(t)$. In particular, for all times $t \in \mathcal{T}$, and features, $k \in \mathcal{K}$, let $\mathbf{x}_k^t$ denote the 2D coordinates of the feature k at t. Then, $\mathbf{\Pi}\left(\mathbf{x}_k^t, d(x_k^t)\right) = \mathbf{X}_k^t$, (3.5), are the 3D coordinates corresponding to $\mathbf{x}_k^t$ and $\mathcal{X}(t)$ is the set of 3D coordinates for all $k \in \mathcal{K}$ at time t. Similarly, a discrete 3D deformation field, $\mathbf{V}(\mathbf{X}, t)$, is computed, relating the 3D feature points at the initial frame, $\mathcal{X}(t^*)$, and another frame, $\mathcal{X}(t)$, as

$$\mathbf{V}\left(\mathbf{X}_k^{t^*}, t\right) = \mathbf{X}_k^t - \mathbf{X}_k^{t^*}. \tag{3.12}$$

A 2D projection of a sample 3D deformation field is shown in Figure 3.7b and 3.7c.

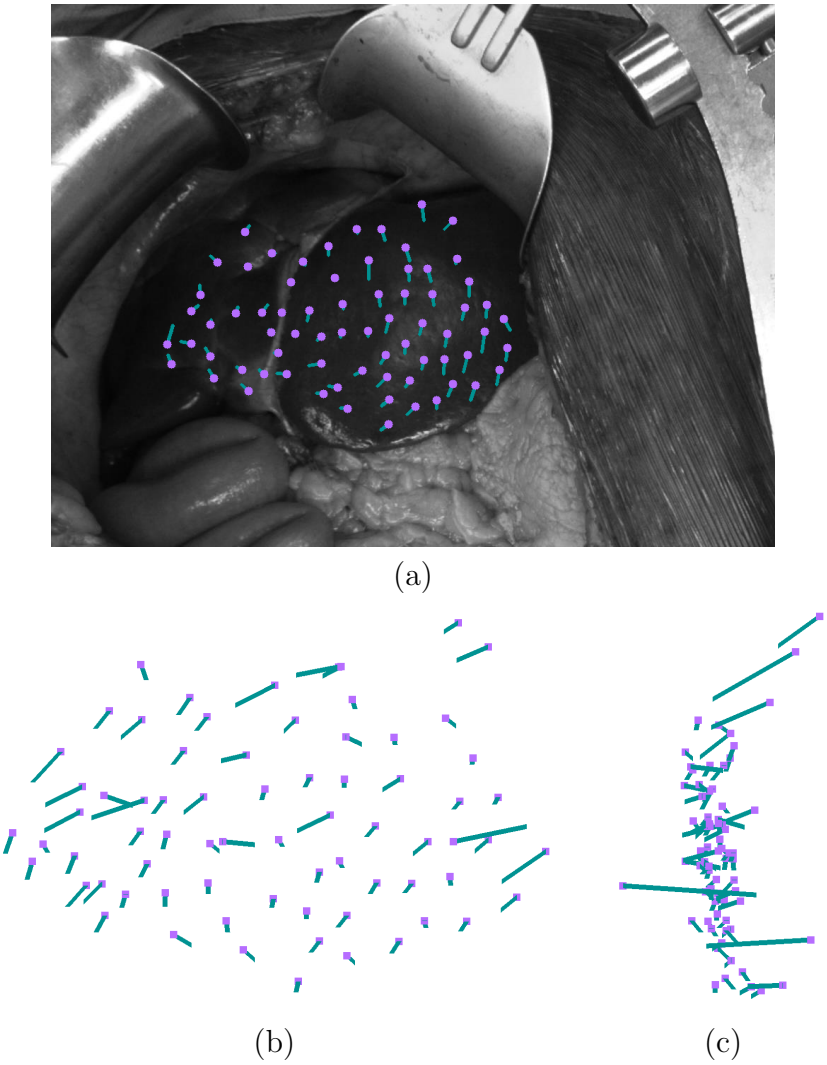


Figure 3.7: 2D Feature Tracking and 3D Deformation Field. (a) 2D "good features to track" (shown as dots) as well as 2D vectors (shown as lines emanating from the dots) showing relative displacements between the current frame and the initial frame are shown for a number of tracked features on the surface of the liver. These 2D features are merged with the disparity information and projected into 3D. Vectors connecting the 3D location of each feature at the initial frame with its current location constitute the 3D deformation field. The deformation field is shown as a 2D projection with a frontal view (b) and side view (c). Similar to (a), in (b) and (c), the dots are the 3D projections of the tracked locations of the "good features to track" extracted at the initial frame at a later frame in the sequence. The lines show the 3D deformation field relating the point sets at the current frame and the initial frame. For clarity in visualization, actual tracked points have been subsampled by a factor of ≈ 2 for display.

3.5 6DOF motion estimation and intraoperative alignment

To perform the registration across frames of the 3D surface reconstructions, a six-*DOF* motion between each frame and the initial frame is recovered based on the previously recovered 3D deformation fields $\mathbf{V}$ as defined in (3.12). The 3D deformation field captures aspects of both the rigid and non-rigid motion of the surface of the liver. Here, we concentrate on exploiting the rigid component; nevertheless, the non-rigid deformation field, $\mathbf{V}$, could be useful in future extensions that aim to model corresponding organ deformations; see Chapter 5.

As the motion of the liver during intraoperative planning/exploratory phases is predominantly rigid [173], the system currently recovers a rigid transformation from the 3D deformation field. The rigid transformation is recovered by a more robust version of an earlier algorithm [3] according to Procrustes

$$(\mathsf{R}, \mathbf{T}) = \underset{\mathsf{R}\in SO(3),\, \mathbf{T}\in\mathbb{R}^3}{\arg\min} \sum_{i=1}^{n} \|(\mathsf{R}\mathbf{P}_i + \mathbf{T}) - \mathbf{Q}_i\|^2 \tag{3.13}$$

where $\mathbf{P}_i$ and $\mathbf{Q}_i$ are 3D points in two surface reconstructions that have been brought into correspondence at n locations tracked by the 3D deformation field, R is a 3×3 rotation matrix, and $\mathbf{T}$ is a 3×1 translation vector. Before computing the rotation, the centroids of the point clouds are computed

$$\bar{\boldsymbol{P}} = \frac{\sum_{i=i}^{n} \boldsymbol{P}_i}{n}, \quad \bar{\boldsymbol{Q}} = \frac{\sum_{i=i}^{n} \boldsymbol{Q}_i}{n} \tag{3.14}$$

and new versions of $\boldsymbol{P}_i$ and $\boldsymbol{Q}_i$ are created that are centred at the origin

$$\boldsymbol{P}'_i = \boldsymbol{P}_i - \bar{\boldsymbol{P}}, \quad \boldsymbol{Q}'_i = \boldsymbol{Q}_i - \bar{\boldsymbol{Q}}, \quad i = 1, 2, \ldots, n. \tag{3.15}$$

A 4×4 covariance matrix, S is then computed as

$$\mathsf{S} = \mathsf{PIQ}^\top, \tag{3.16}$$

where P and Q are the $4 \times n$ matrices that have $\boldsymbol{P}'_i$ and $\boldsymbol{Q}'_i$ as their columns, respectively, and I is an $n \times n$ identity matrix. The *Singular Value Decomposition (SVD)* decomposition of the covariance matrix,

$$\mathsf{S} = \mathsf{U\Sigma V}^\top, \tag{3.17}$$

is used to compute the rotation as

$$\mathsf{R} = \mathsf{V} \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ & & & & \det(\mathsf{VU}^\top) \end{bmatrix} \mathsf{U}^\top. \tag{3.18}$$

The optimal translation is then the residual created in the origin-centric point clouds after the rotation is applied

$$\mathbf{T} = \bar{\boldsymbol{Q}} - \mathsf{R}\bar{\boldsymbol{P}}. \tag{3.19}$$

The solution is made more robust by using a modified version of *Random Sample*

Consensus (RANSAC) [48] to minimize the effect of outlier correspondences, poor depth estimates, or those most effected by any non-rigidity in the viewed scene. Traditional *RANSAC* is shown in Algorithm 4 with the modifications being shown in Algorithm 5. The modified approach differs from traditional *RANSAC* in two main ways. First, *RANSAC* traditionally uses the minimum number of point correspondences necessary to estimate a 3D transformation, i.e., three. The modified version of the algorithm uses a user-defined number of point correspondences, p, that is set to 30 point pairs during the evaluation process. Second, the traditional *RANSAC* approach terminates once the application of the motion estimate results in the fraction of inliers to the total number of points exceeding some manually defined threshold, τ. The modified version of the algorithm runs for a set number of iterations returning the motion estimate resulting in the largest number of inliers. These modifications resulted in stable motion estimates being produced across all intraoperative and laboratory datasets with common parameters being used across datasets.

Notably, registering incoming frames back to the initial frame ameliorates issues of registration drift that can occur in the most prevalent alternative approaches of chaining sequential shape-based registrations between adjacent frames (i.e., *ICP*) over a long sequence. An example final alignment between a 3D surface reconstruction and volumetric model is shown in Figure 3.8.

Algorithm 4: Random Sample Consensus (RANSAC) for Rigid Motion Estimation

1: Select three pairs of 3D points at random from the two sets of point correspondences relating the two frames.
2: Estimate a rigid rotation, R, and translation, $\mathbf{T}$, from the three selected pairs of points.
3: Determine how many points from the two sets of point correspondences, after applying the rigid rotation and translation estimate, fall within a predefined distance threshold, ϵ. These will be the model *inliers*.
4: If the fraction of the number of inliers over the total number of point correspondences exceeds a predefined threshold, τ, re-estimate the rigid motion estimate using all the inliers and terminate (returning the newly calculated R and $\mathbf{T}$).
5: Otherwise, repeat steps 1 through 4 (maximum N times).

Algorithm 5: Modified Random Sample Consensus (RANSAC) for Rigid Motion Estimation

1: Select p pairs of 3D points at random from the two sets of point correspondences relating the two frames.
2: Estimate an optimal rigid rotation, R, and translation, $\mathbf{T}$, from the p selected pairs of points.
3: Determine how many points from the two sets of point correspondences, after applying the rigid rotation and translation estimate, fall within a predefined distance threshold, ϵ. These will be the model *inliers*.
4: If the number of inliers is greater than the maximum number of inliers previously seen or if this is the first iteration, set the maximum to the current number of inliers and update the transformation estimate to the newly calculated R and $\mathbf{T}$.
5: Repeat steps 1 through 4 (N times).

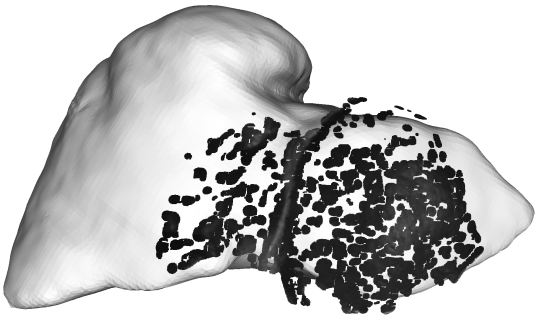


Figure 3.8: Intraoperative Alignment. The darker points in the foreground represent the surface reconstruction of the visible portion of the patient's liver generated by the stereo reconstruction algorithm and subsequent postprocessing, as described in Section 3.3.2. The intensity of a given point matches the grayscale value of the corresponding pixel in the reference (right) image. The brighter surface model in the background is the preoperative *CT* scan of the same patient's liver. The initial alignment of these two data sources was performed manually, in accordance with the description provided in Section 3.2.2.

3.6 Implementation

The described system has been implemented in software. Implementation details for all aspects of the system, and all parameter values, are provided in Appendix A. The most time critical components of the system, when considering potential integration into surgical workflow, are the module generating dense disparity information from the stereo imagery, the module that maintains the 2D feature tracks, and the module that combines these pieces of information into a 3D deformation field and computes the rigid registration estimate relating the incoming frame back to the initial frame of the sequence. The system runtime was measured by independently measuring the runtime of these three components and concatenating them so as to run fully sequentially (i.e., no concurrency at the pipeline level). Following this methodology, given input 640×480 images, the core system components execute at 8 Frames Per Second (FPS) on an Intel Xeon CPU E5-1650 v4 3.6GHz processor with 256GB of RAM. Considerable speed-up is anticipated with a *Graphics Processing*

Unit (GPU) implementation, e.g., the slowest component of the system is stereo correspondence, which already has been previously ported to run on a 240 core *GPU* (nVIDIA GeForce GTX285) with 640×480 images and 256 disparity levels at 100 *Frames Per Second (FPS)* [178]. Therefore, there is overall potential for the system, minus initialization, to maintain an alignment between preoperative and intraoperative data at rates greater than the frame rate of typical input video sequences. For reference, the semi-automated initialization strategy, which runs outside of the main loop of the system, took ≈ 33 minutes (without parallelization) using Algorithm 2 to generate the 50 registration candidates used for evaluating ***Intraoperative Case 1***.

3.7 Summary

This chapter presented a novel end-to-end system for fast surface reconstruction and motion estimation for alignment with a preoperative *CT* scan targeted to open liver resection tasks. The pipeline employs two different initialization techniques: a traditional manual 3D-3D surface registration of the 3D reconstruction at the initial frame of the sequence to the preoperative *CT* scan and a semi-automated technique that uses an efficient *BnB* strategy to find a global registration [208]. A hand delineation of pixels belonging to the liver in the right image of the initial frame is also provided as part of the system initialization in the form of an image mask.

By exploiting the hand provided mask, structure and motion belonging to the liver can be successfully decoupled from the surgical scene as a whole (which behaves in a very deformable manner as opposed to the predominantly rigid behaviour of

the liver undergoing respiratory motion [173]). 3D structure of the liver is extracted through the use of an adaptive *CTF* stereo matching technique [179]. This approach provides a fast estimate of the depth of points belonging to the liver surface, while helping to preserve 3D discontinuities in an attempt to maximize the amount of the liver surface recovered.

"Good features to track" [168] are extracted from the liver surface at the initial frame and tracked throughout the video sequence. The tracked features provide an estimate of correspondences to visible features on the liver surface over time. By combining the tracked 2D features with the recovered 3D surface, a 3D deformation field is produced allowing for any frame in the image sequence to have a set of 3D feature correspondences back to the initial frame. A rigid motion estimate is computed from the non-rigid deformation field (via a robust version of an earlier approach [3]) and combined with the initial hand registration of the initial frame to register subsequent frames back to the preoperative *CT* scan.

Chapter 4

Empirical evaluation

The system detailed in Chapter 3 has been evaluated empirically in both controlled laboratory conditions and with clinical data acquired during liver resection tasks conducted at *Memorial Sloan Kettering Cancer Center (MSKCC)*. Laboratory testing allowed for evaluation of both the comparison of *Surface Registration Error (SRE)* as well as comparison of the motion estimation required for optical organ registration tasks (typically a level of analysis not available in an intraoperative setting).

This chapter is laid out as follows. Section 4.1 provides details of the imaging setup used for both the laboratory and intraoperative test conditions. Section 4.2 provides details about the laboratory and intraoperative datasets that were used to evaluate the developed *SN* system. Section 4.3 provides comparative evaluation of the manual and proposed semi-automated registration technique. Section 4.4 provides estimates of the error in the motion profiles recovered by the system in the laboratory datasets (where ground-truth motion estimates are present). Section 4.5 uses *Surface Registration Error (SRE)* to provide a performance metric for how

close the reconstructed and registered intraoperative surface estimate matches the corresponding surface of the preoperatively acquired *CT* scan of the liver. Discussion of the results presented in Sections 4.3 - 4.5 is deferred to Section 4.6.

4.1 Image capture

All datasets were acquired using two Grasshopper2 GS2-GE-20S4C GigE machine vision cameras[1] with a baseline of ≈ 60 mm at a targeted standoff distance of ≈ 700 mm (see Figure 4.1). The Grasshopper2 GS2-GE-20S4C GigE machine vision cameras provide a $1/1.8''$ *Charge-Coupled Device (CCD)* sensor with a maximum resolution of 1624×1224 (a pixel size of $4.4\,\mu$m). Two C-mount lenses[2] (maximum aperture: f/1.4, focal length: 8 mm and lens diameter (ø): 25.5 mm) were used. Synchronized stereo images were captured at a resolution of 1024×768 in grayscale at a targeted frame rate of 30 *FPS*. All intraoperative evaluations take place in open liver resection conditions. Laboratory evaluations are meant to replicate the relative positioning of the liver and camera setup as in ideal open surgical scenarios. Regarding the illumination, we make use of two soft boxes, providing a diffuse lighting source, are placed on either side of the liver phantom (approximately 1 m on each side of the target) for the laboratory image sequences. For the intraoperative sequences, we make use of whatever lighting was available in the surgical theatre. No special accommodations were made. Figure 4.2 shows pictures of the test scenarios as well as acquired images for both example laboratory and operating room conditions.

[1] Grasshopper2 GigE cameras were manufactured by Point Grey Research, Vancouver, BC, Canada (now named FLIR).

[2] Lenses were manufactured by Goyo Optical, Japan.

Figure 4.1: Stereo Sensor. Two calibrated video cameras are mounted in a metal enclosure. See text for details.

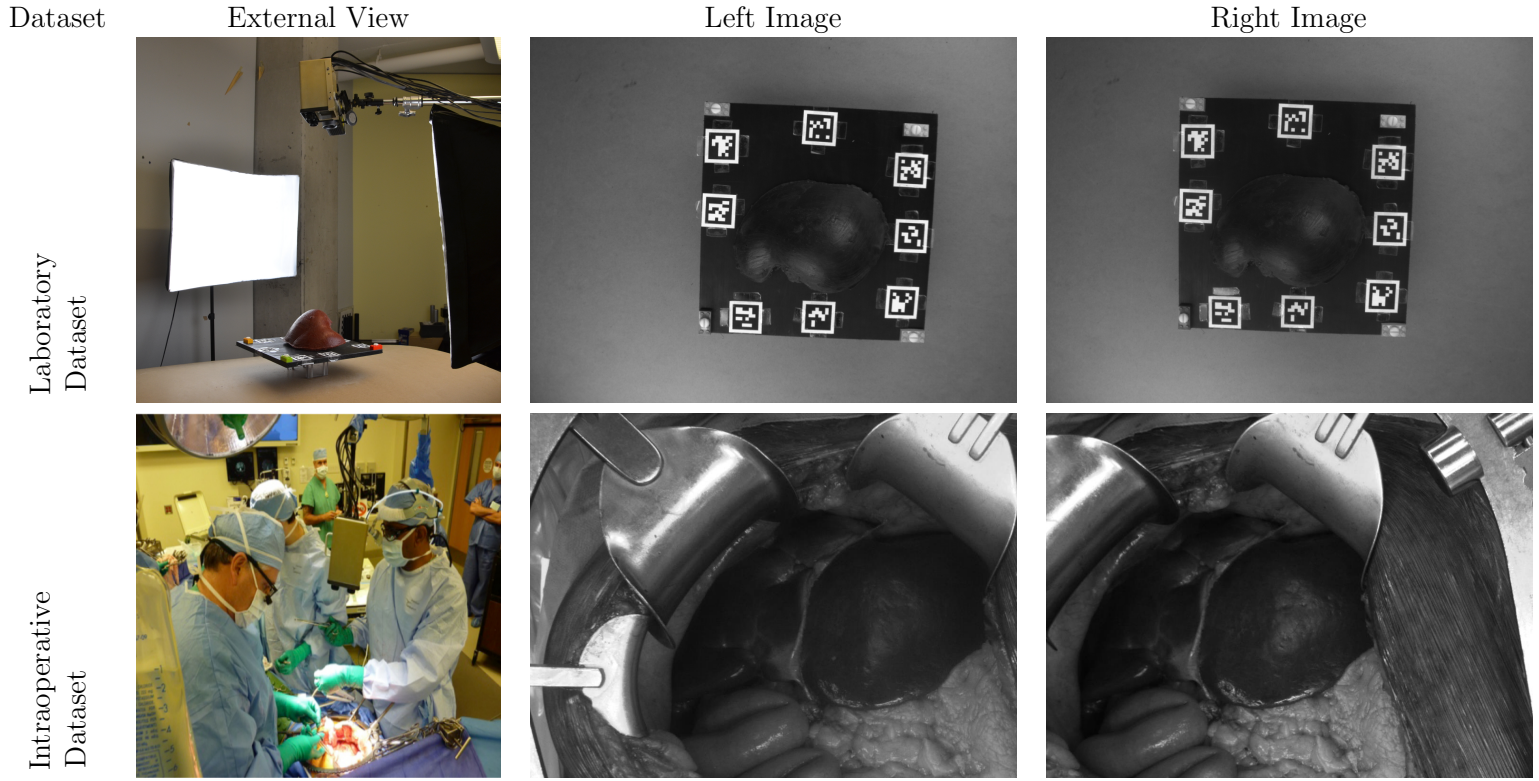


Figure 4.2: Example Image Capture in Both Laboratory and Surgical Settings.

4.2 Datasets

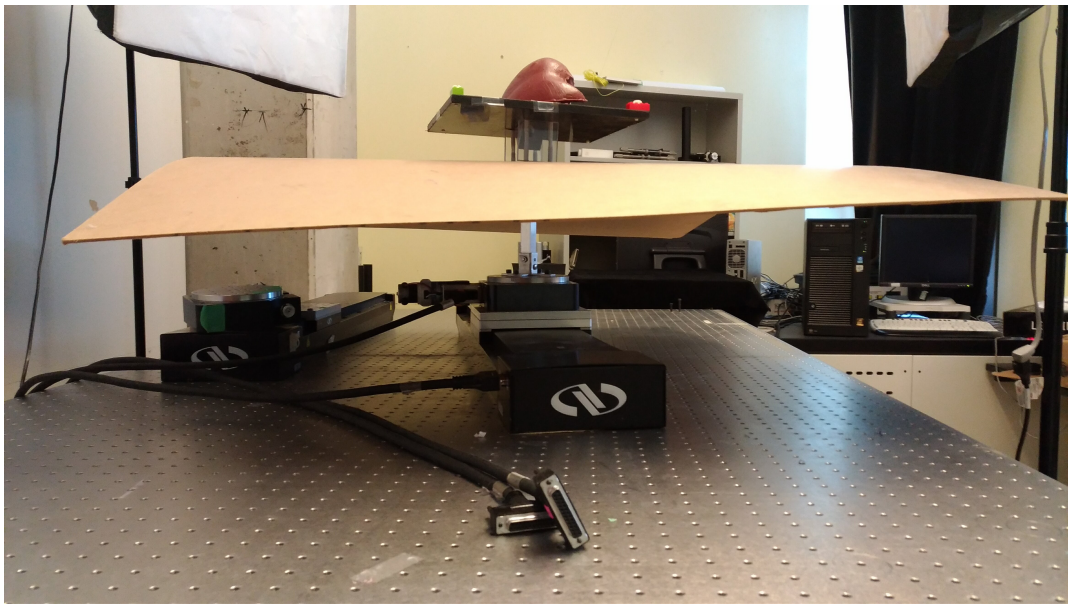
The laboratory dataset consists of three calibrated stereo image sequences of a silicone liver phantom attached to a motion control platform (one linear and one rotational stage)[3] for precise motion control; see Figure 4.3. The phantom is rigidly affixed to a fiducial board containing 2D *AR*-tag fiducials [51] located on the outer perimeter of the fiducial board as well as 3D LEGO[4] fiducials located in the corners of the fiducial board. These fiducials are used to aid in the initial manual alignment of the reconstructed point cloud to the corresponding *CT* scan and for verification of the provided motion commands. Each laboratory test condition consists of a 31 frame sequence with evenly spaced samples along the motion trajectory. Three different motion profiles were tested: ***Translation-Only*** (1 mm increments), ***Rotation-Only*** (1 degree increments), and ***Translation+Rotation*** motions superimposed. As mentioned previously, the largest component of motion in open liver surgery conditions is due to respiration causing large translational motion in the order of $\approx 10-75$ mm [33]. The laboratory dataset covers average translational motions, in the range of $0-30$ mm, and greater than average rotational motions, in the range of $0-30$ degrees, with motion increments being chosen in accordance with the known precision of the linear and rotational Newport motion stages. A representative frame is shown in this chapter in Figure 4.4. Additional frames are shown in Appendix B, Figure B.1. The laboratory dataset image sequences were acquired at an approximate standoff distance of 700 mm to the liver phantom (equivalent to the targeted distance for the device when used intraoperatively).

[3]The motion control platform and motion stages are manufactured by Newport Corporation, Irvine, CA, United States.

[4]LEGO bricks are a play material manufactured by The LEGO Group, Billund, Denmark.

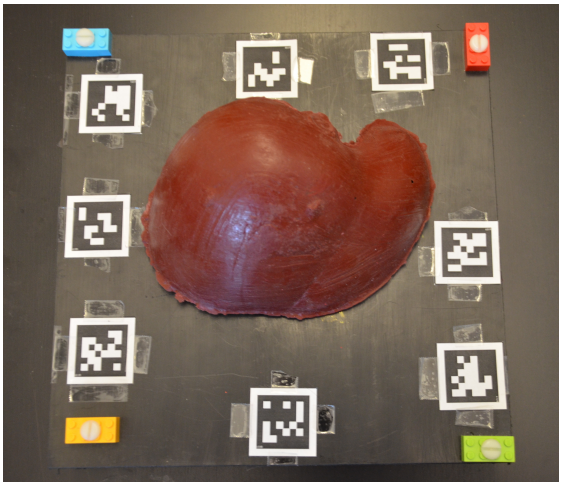
Three intraoperative data sequences were captured during three separate surgical interventions requiring open liver resection at *MSKCC*. The target standoff distance for each surgery was approximately 700 mm. The first intraoperative dataset, ***Intraoperative Case 1***, consists of 350 frames. The second and third intraoperative datasets, ***Intraoperative Case 2*** and ***Intraoperative Case 3***, are considerably longer in duration consisting of 1729 and 1726 frames, respectively. Representative frames are shown here in Figure 4.4. Additional frames are shown in Appendix B, Figures B.2-B.4. Intraoperative datasets were captured with a frame rate of 30 *FPS*.

In both the laboratory and operating room, the same calibrated stereo video camera, as specified in Section 4.1, was used for image acquisition (the gold box in Figure 4.1 and in the upper middle portion of the external views shown in Figures 4.2 and 4.3b).



(a)

(b)



(c)

Figure 4.3: Laboratory Dataset Experimental Setup. Figure (a) depicts the general experimental setup where the linear (bottom) and rotation (top) stages of the motion platform are visible as well as the linkages connecting the stages directly to the liver phantom and fiducial board. Figure (b) shows the same setup (including illumination, plywood backdrop, liver phantom and fiducial board as well as the stereo camera positioned above the target liver phantom). Figure (c) shows an overhead close-up of the fiducial board and liver phantom.

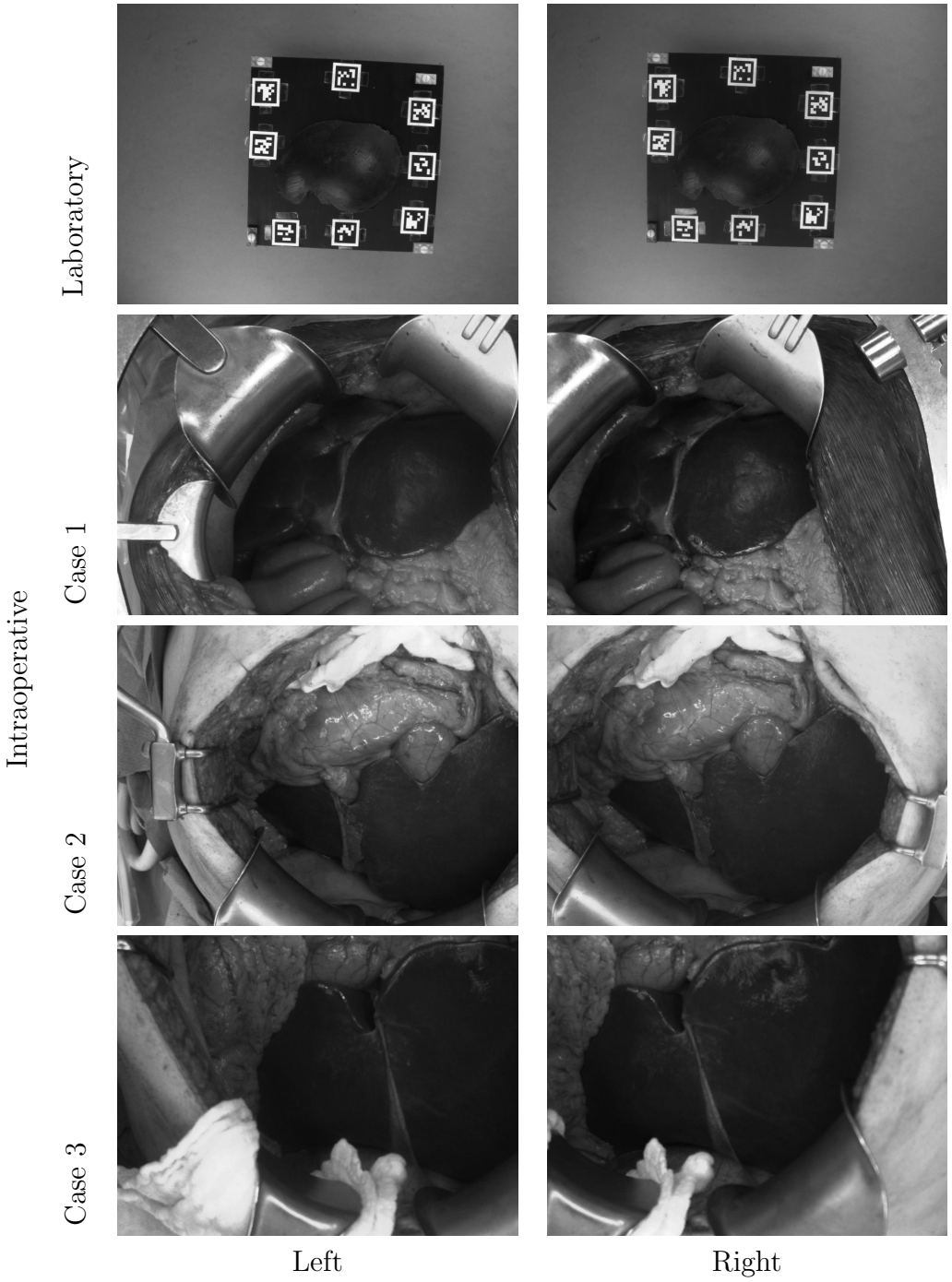


Figure 4.4: Dataset Overview. An example stereo pair from each of the employed datasets is shown. For simplicity, a single left-right view representative of the appearance of all three conditions in the laboratory dataset is shown (as the three datasets only differ by the relative motion of the phantom liver and the camera).

4.3 Initialization method comparison

As mentioned previously in Section 3.2, two initialization methods were developed for use in this system. First, a standard approach involving manual registration of the partial 3D stereo surface reconstruction at the initial frame to the preoperative *CT* scan was employed. This method for initialization is analogous to the standard method employed for the vast majority of systems in the *SN* community. The second method employs a modified version of *Globally Optimal Iterative Closest Point (GO-ICP)* to provide registration candidates to a human-in-the-loop. This section evaluates the utility of the *GO-ICP* algorithm for initialization of both the laboratory and intraoperative datasets. First, the ability of the *GO-ICP* algorithm to find a registration candidate that is visually similar to the hand registration within a nominal number of candidates is explored. Second, registration candidates resulting from the *GO-ICP* algorithm are quantitatively evaluated and compared to the hand registration candidate.

4.3.1 Finding a registration candidate

Standard *GO-ICP* attempts to find a globally optimal rigid registration using a *BnB* strategy to search the entire 3D motion space, $SE(3)$. The modified version, presented in Section 3.2.3, is able to produce the n-best registration candidates in the search space. For the purposes of the experiments in this dissertation the 50 best registration candidates between the 3D stereo surface reconstruction of the initial frame and the preoperative *CT* scan are generated from the modified *GO-ICP* algorithm to see if a meaningful registration candidate is produced (determined via visual inspection). Table 4.1 and Figures 4.5-4.10 show the results of this process.

Interestingly, all of the laboratory datasets showed a visually meaningful registration result in the best *GO-ICP* registration candidate. For the three intraoperative cases, only ***Intraoperative Case 1*** exhibited a visually meaningful registration candidate within the 50 best candidates (and did so at registration candidate 14/50). An anterior view of the first 20 registration candidates for this case are shown in Figure 4.11.

Dataset	Condition	Acceptable Candidate in Top 50	First Acceptable Position
Laboratory	Translation-Only	Yes	1
	Rotation-Only	Yes	1
	Translation+Rotation	Yes	1
Intraoperative	Case 1	Yes	14
	Case 2	No	N/A
	Case 3	No	N/A

Table 4.1: Globally Optimal Iterative Closest Point (GO-ICP) Candidate Search Success Statistics.

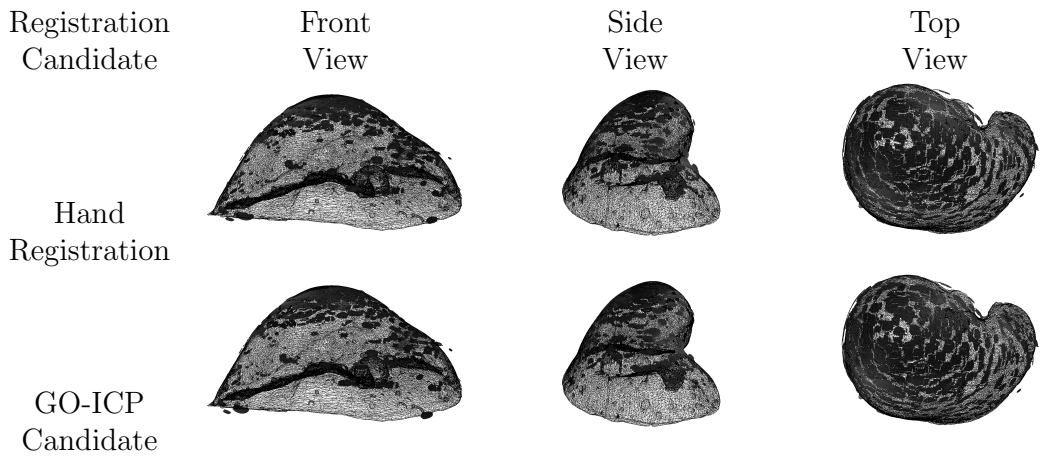


Figure 4.5: Registration Initialization Candidates - Laboratory Dataset: Translation-Only. Darkest points correspond to the recovered surface model, as overlaid on a fine mesh of the volumetric model.

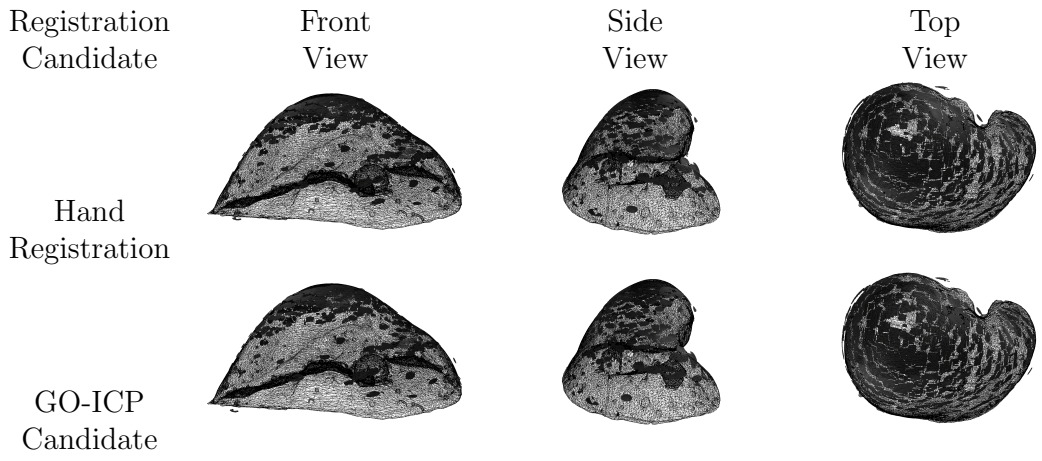


Figure 4.6: Registration Initialization Candidates - Laboratory Dataset: Rotation-Only. Visualization as in Fig. 4.5.

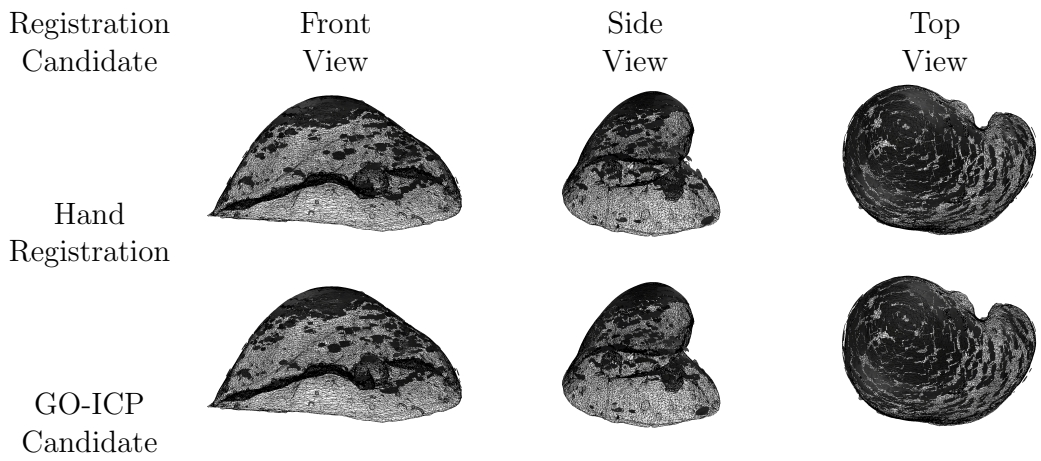


Figure 4.7: Registration Initialization Candidates - Laboratory Dataset: Translation+Rotation. Visualization as in Fig. 4.5.

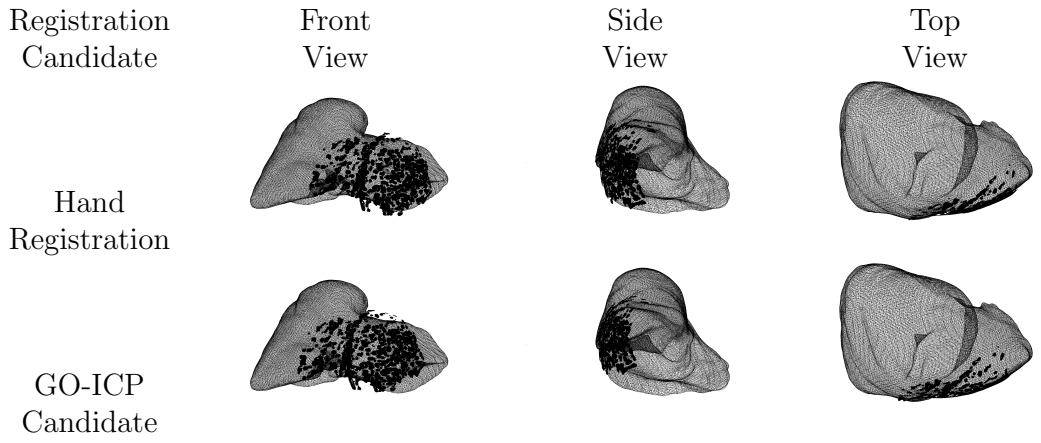


Figure 4.8: Registration Initialization Candidates - Intraoperative Case 1. Visualization as in Fig. 4.5.

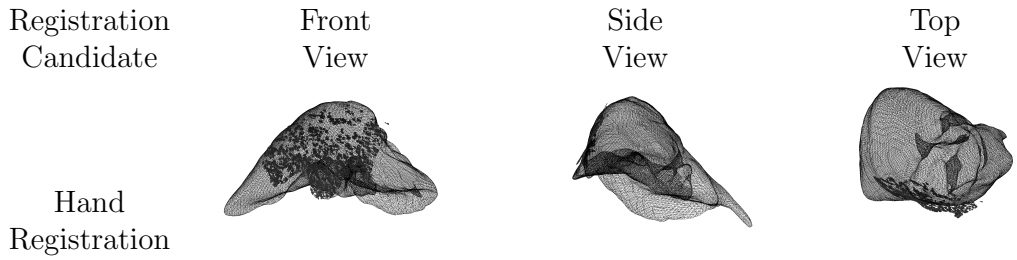


Figure 4.9: Registration Initialization Candidates - Intraoperative Case 2. No acceptable *GO-ICP* result was returned for this case. Visualization as in Fig. 4.5.

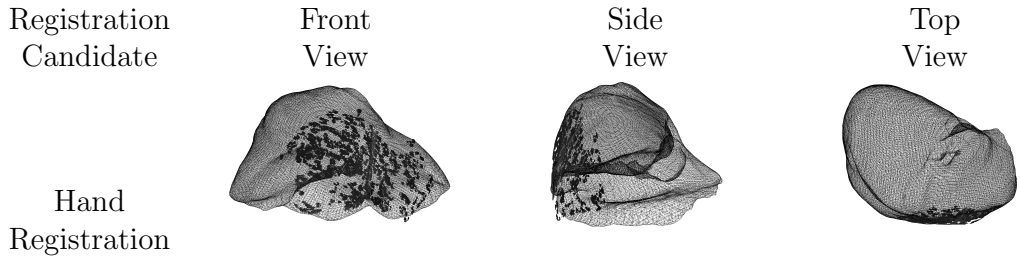


Figure 4.10: Registration Initialization Candidates - Intraoperative Case 3. No acceptable *GO-ICP* result was returned for this case. Visualization as in Fig. 4.5.

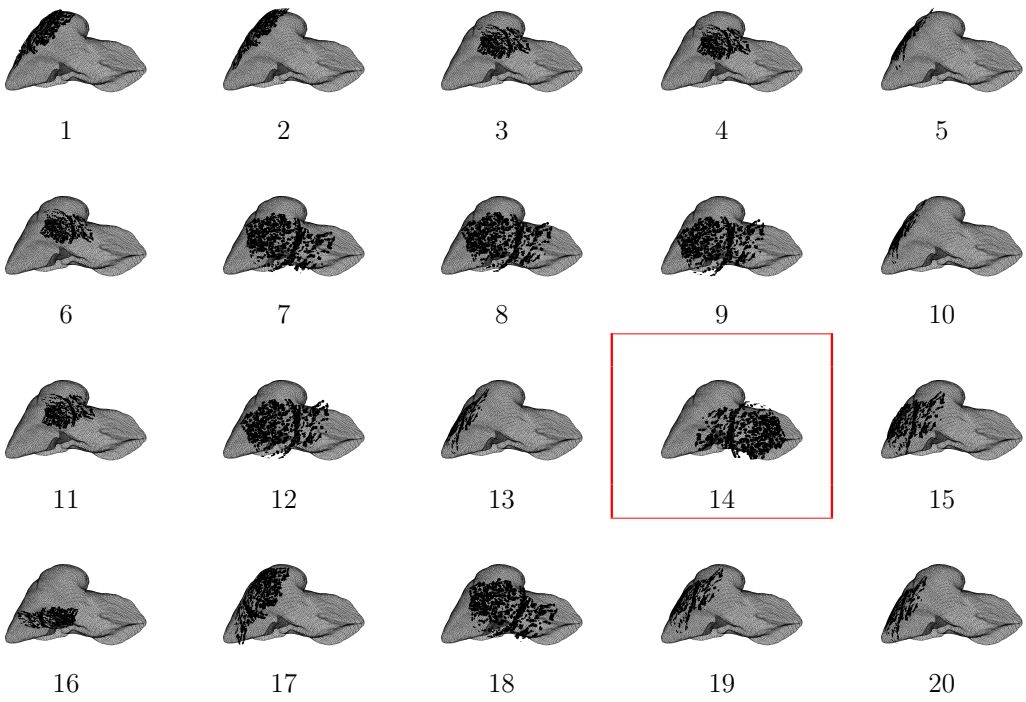


Figure 4.11: Globally Optimal Iterative Closest Point (GO-ICP) Registration Initialization Candidates - Intraoperative Case 1. The first 20 registration candidates output by the modified version of *GO-ICP* are shown for ***Intraoperative Case 1***. The registration candidate chosen by the human-in-the-loop, candidate 14, is highlighted in red. Visualization as in Fig. 4.5.

4.3.2 Manual versus GO-ICP registration candidates

Once a visually convincing (qualitative) registration candidate has been found further analysis of the registration candidates was needed to assess the viability of the semi-automated registration technique. As such, a measure of *Surface Registration Error (SRE)*,

$$\text{SRE}_i = \|\mathbf{P}_i - \mathbf{P}_i^*\|, \tag{4.1}$$

where $\mathbf{P}_i$ is a point in the registered stereo reconstruction of the surface and $\mathbf{P}_i^*$ is the closest point to $\mathbf{P}_i$ in the preoperative volume after registration, was calculated for each of the registration candidates.

The results of this analysis are reported in three ways. First, the error distribution for each registration candidate is plotted in the form of a box plot in Figure 4.12. Second, the mean and standard deviation statistics for the registration candidates are enumerated in Table 4.3.2. Third, the *SRE* is depicted as a colour-coded 3D surface model in Figures 4.14 and 4.15 showing the spatial distribution of error over the surface reconstruction.

Dataset	Condition	Manual Registration		GO-ICP Registration	
		SRE (mean)	SRE (std)	SRE (mean)	SRE (std)
Laboratory	Rotation-Only	0.8968	0.8175	0.9082	0.7755
	Translation-Only	0.8996	0.8067	0.9008	0.7486
	Translation+Rotation	0.7951	0.7433	0.8219	0.6802
Intraoperative	Case 1	1.3652	1.4115	1.7935	1.8967
	Case 2	2.8479	2.9421	N/A	N/A
	Case 3	2.5257	1.4280	N/A	N/A

Table 4.2: Mean and Standard Deviation of Initialization Candidates. All measurements are provided in mm.

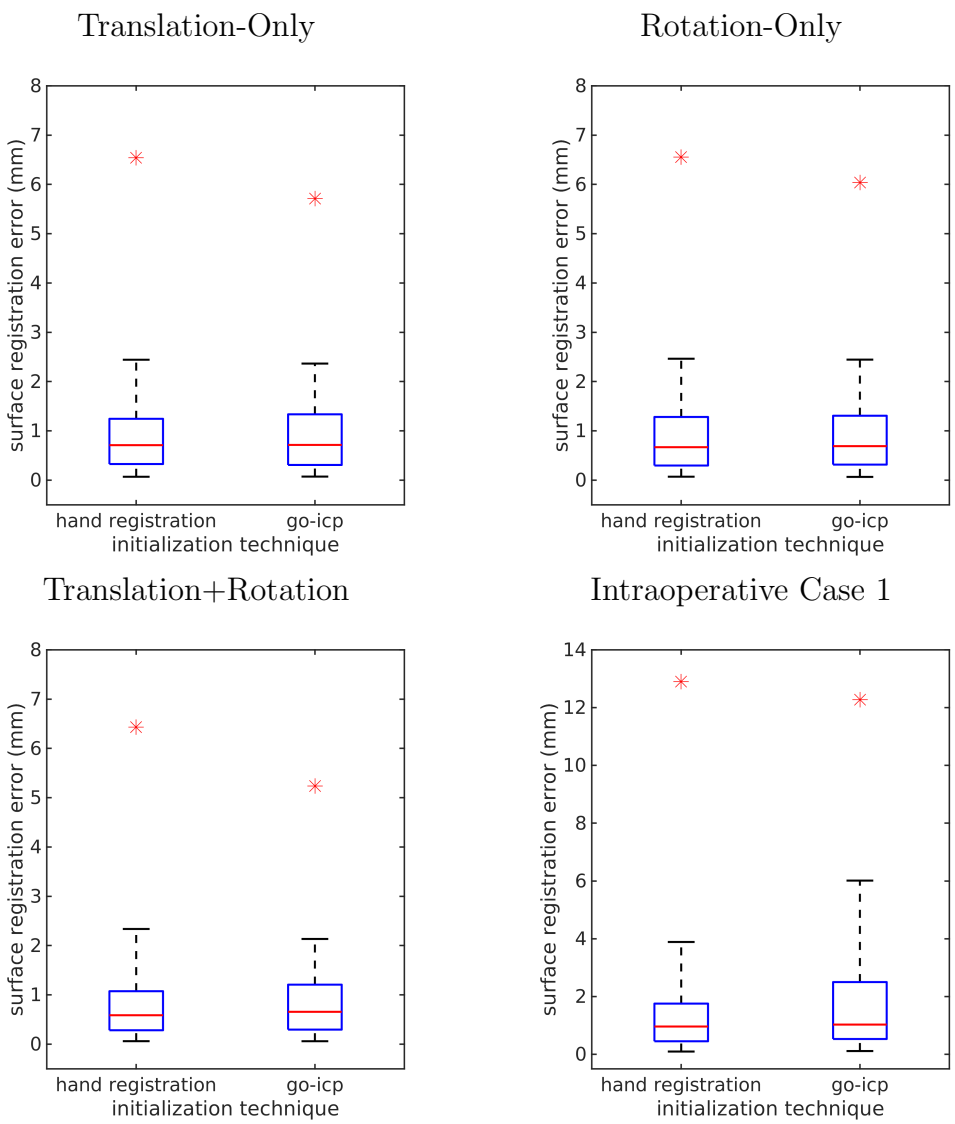


Figure 4.12: Globally Optimal Iterative Closest Point (GO-ICP) vs. Hand Initialization Surface Registration Error (SRE). The *SRE* for the hand registration candidate and the best candidate resulting from the use of the modified *GO-ICP* algorithm are shown as box plots. Blue boxes cover the interquartile range, with the red line showing the median value. Whiskers cover 90% of the reconstructed point clouds. The largest outlier for each registration candidate is indicated by a red asterisk. Note that the y-axis on the three laboratory dataset conditions range from 0-8 mm. The y-axis on the intraoperative case ranges from 0-14 mm.

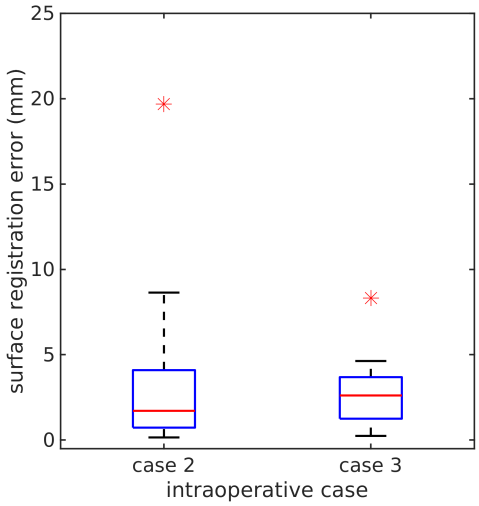


Figure 4.13: Hand Initialization Surface Registration Error (SRE) for Intraoperative Case 2 and Intraoperative Case 3. The *SRE* for the hand registration candidate on the two unsuccessful *GO-ICP* cases, ***Intraoperative Case 2*** and ***Intraoperative Case 3*** are shown as box plots. Blue boxes cover the interquartile range, with the red line showing the median value. Whiskers cover 90% of the reconstructed point clouds. The largest outlier for each registration candidate is indicated by a red asterisk. No acceptable semi-automated registrations were produced for these cases.

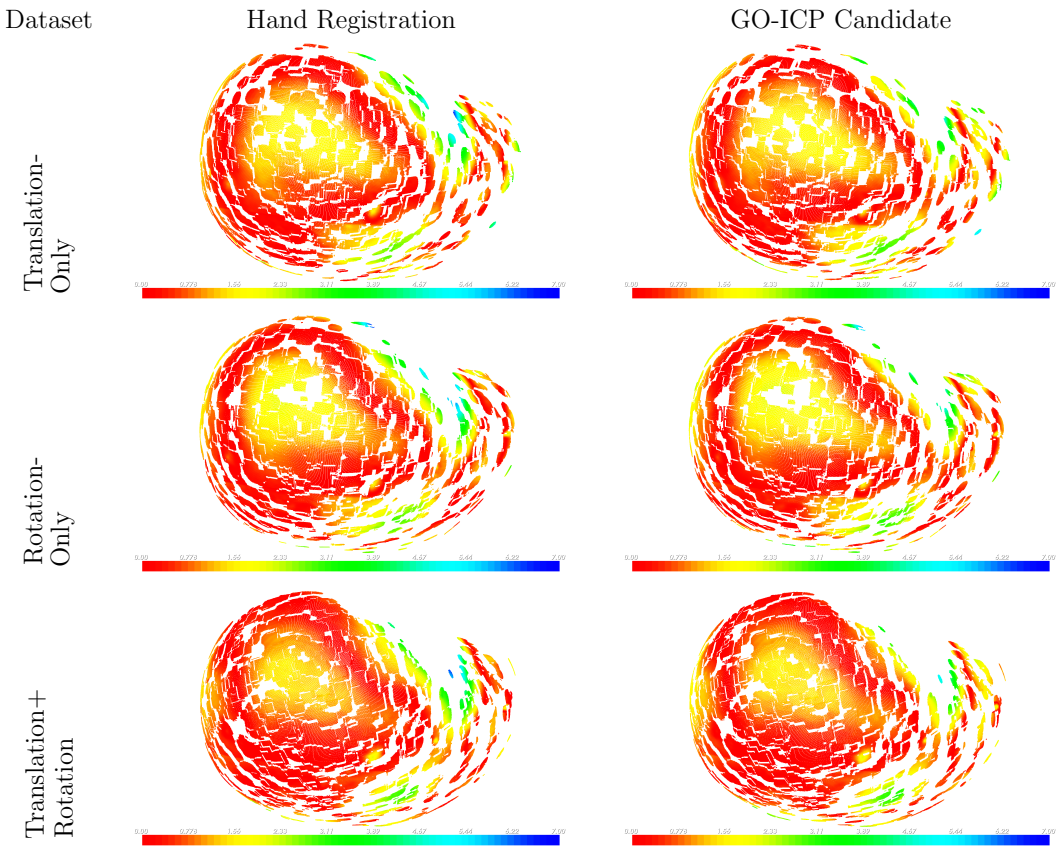


Figure 4.14: Globally Optimal Iterative Closest Point (GO-ICP) vs. Hand Initialization Registration Candidates - Laboratory Dataset. The *Surface Registration Error (SRE)* distributions of the hand registration candidate and the leading *GO-ICP* registration candidate for the three laboratory dataset test conditions (***Translation-Only***, ***Rotation-Only*** and ***Translation+Rotation***) are shown. Errors are depicted with respect to the location and magnitude of the error on the reconstructed surface. The visual scale for the registration errors are shown in the bars below each scan and are the same across all cases (ranging from 0 mm - pure red, to 7 mm - pure blue).

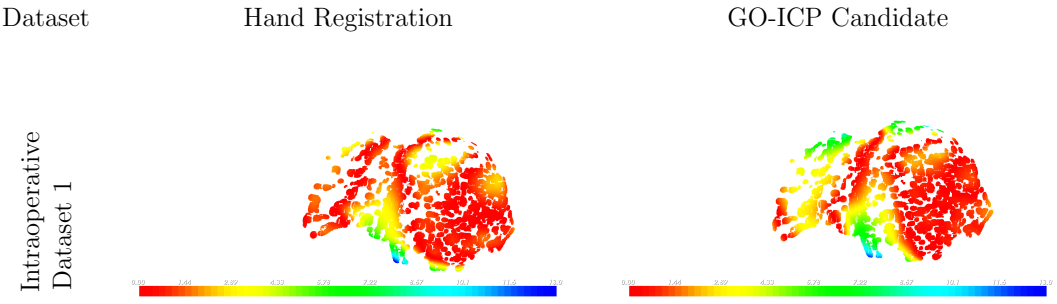


Figure 4.15: Globally Optimal Iterative Closest Point (GO-ICP) vs. Hand Initialization Registration Candidates - Intraoperative Case 1. The *Surface Registration Error (SRE)* distributions of the hand registration candidate and the leading *GO-ICP* registration candidate for ***Intraoperative Case 1*** is shown. Errors are depicted with respect to the location and magnitude of the error on the reconstructed surface. The visual scale for the registration errors are shown in the bars below both scans and are the same for both cases (ranging from 0 mm - pure red, to 13 mm - pure blue).

4.4 Motion estimation

The translation (magnitude) and rotation components of the recovered motions are reported for the laboratory datasets in Figure 4.16. (Recall that the intraoperative sequences did not support ground-truth motion; so, no quantitative motion estimates are provided.) For each of the three tested motion profiles (***Translation-Only***, ***Rotation-Only***, and ***Translation+Rotation***), the recovered translation magnitude is shown in panels (a, c, and e), while the recovered rotation angles are shown in panels (b, d, and f). At any given frame, motion was recovered with respect to the initial frame in the sequence. For ***Rotation-Only*** it is seen that the recovered angle accurately tracks the true motion (interframe increments of 1 degree), approximately a line of slope one and the recovered translation is correctly very small. (Note that since the true translation has zero magnitude, the error between the recovered and ground-truth is the same as the recovered and only the

recovered is shown.) For the ***Translation-Only*** case, the results are exactly complementary to those of ***Rotation-Only***, again showing very accurate performance. Similarly, the ***Translation+Rotation*** case shows the desired combination of the other two cases. Notably, motion drift is not affecting the interframe estimation over these sequences owing to the motion always being computed relative to the initial frame. In contrast, typical approaches that chain transformation estimates between adjacent frames to relate a given frame back to a keyframe would be susceptible to drift.

Table 4.3 provides a summary of the mean and standard deviation of the reported motion errors for the three laboratory datasets. Note that mean errors and standard deviation of the motion estimates are not provided for the intraoperative datasets as no ground-truth motion information is available.

	Translation Magnitude (mm)		Rotation Angle (deg)	
Dataset	mean	std	mean	std
Rotation-Only	1.8975	0.6547	-0.4124	0.5126
Translation-Only	-0.1440	0.6560	1.0898	0.5744
Translation+Rotation	0.5654	0.4092	-0.4507	0.4095

Table 4.3: Mean and Standard Deviation of Motion Error.

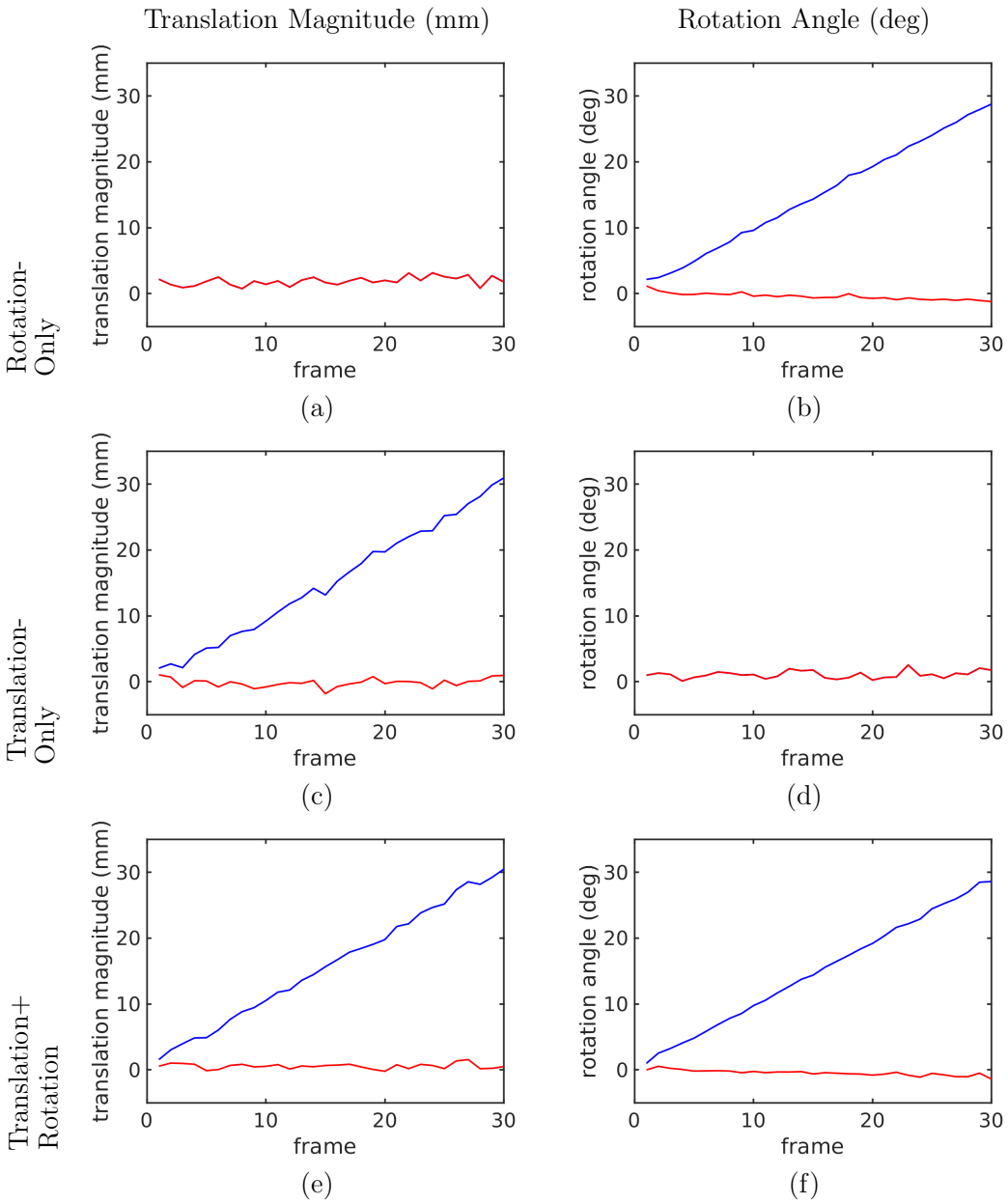


Figure 4.16: Motion Recovery Results. The recovered translation magnitude (a, c, and e) and rotation (b, d, and f) estimates are plotted for each of the three laboratory motion profiles (***Translation-Only***, ***Rotation-Only***, and ***Translation+Rotation***). The blue line indicates the motion estimate; the red line depicts the deviation of the motion estimate from the ground-truth, as actuated by the motion platform. The two plots showing a single curve correspond to measurements where the ground-truth signal was zero motion; hence, the recovered and error values are the same and only the recovered is shown.

4.5 Surface registration error

A measure of *Surface Registration Error (SRE)*, Equation (4.1), was calculated for all laboratory and intraoperative dataset sequences. Figures 4.17-4.22 depict a sample stereo image pair from each of the laboratory and intraoperative dataset conditions along with a visual depiction of the intraoperative surface reconstruction registered to the preoperative *CT* scan at three frames across the sequence. Figures showing all reported frames co-registered, with and without the corresponding *CT* for reference, are shown in Figures C.4-C.9 in Appendix C. The plots in Figures 4.23 and 4.24 show the distribution of *SRE* over all points in the reconstructed point clouds for the selected frames. Registration between *CT* coordinates and the initial frame of each laboratory sequence was performed using the semi-automated initialization technique, as specified in Section 3.2.3. *SRE* results, initialized via identification of five cross-modality features located on the platform supporting the phantom in both the stereo imagery and in the *CT* scan, are provided in Appendix C. Intraoperative data was initialized using the semi-automated initialization technique when a successful candidate was found (i.e., ***Intraoperative Case 1***). Otherwise, a manual registration process, specified in Section 3.2.2, was used. Analogous results using manual initialization on ***Intraoperative Case 1*** are shown in Appendix C. Thus, for all cases where the semi-automated initialization procedure yielded valid registrations, they are used in the results presented here. The corresponding fully manual initialization results are relegated to an appendix for the sake of comparison. For cases where the semi-automated approach failed, fully manual initialization is used in this chapter.

The initialization procedure defines the transformation relating the initial (key)

frame to the *CT* coordinate space. Subsequent frames were registered to the initial frame using the approach described in Section 3.5 and then registered to *CT* space using the transformation relating the initial frame to the *CT* scan. Figure 4.23 shows *SRE* for the three laboratory datasets. Mean errors are approximately 1 mm with over 99% of the points lying under 4 mm *SRE* across the laboratory datasets. Figure 4.24 shows *SRE* for the three intraoperative datasets. For the intraoperative dataset conditions, the errors are larger in general with a larger standard deviation. ***Intraoperative Case 1*** is most similar to the laboratory dataset conditions in terms of *SRE* comparison with an average error of $\approx$1.93 mm. ***Intraoperative Case 2*** and ***Intraoperative Case 3*** both experienced similar average error of $\approx$3.2 mm.

Notably, the interframe registration algorithm never explicitly minimizes *SRE* by applying shape-based registration techniques (e.g., iterative closest point); therefore, the registration is data-driven with respect to the actual positions of identified features on the surface of the organ and is more likely to produce a physically meaningful registration.

Table 4.4 provides a summary of the average and standard deviation of the reported *SRE* for the three laboratory dataset conditions and the three intraoperative dataset conditions.

Similarly to the Middlebury stereo dataset evaluation protocol [165], it is often useful to provide a statistic that represents the portion of the data falling within some range of utility. In the field of stereo correspondence, this often takes the form of the percentage of reported pixels lying within one pixel disparity error from ground-truth. The safety margin around a potential tumour location in laparoscopic interventions is $10 - 25$ mm [140]. Several independent researchers

have proposed an error of < 5 mm as being of interest in *SN* applications to the liver (e.g., [93, 158, 14]). This threshold has been rationalized in a variety of ways: comparison to tolerances in other parts of the body [93], ability to restore intuitive spatial understanding to the surgeon [158], and comparison to limits around typical surgical margins [14]. The < 5 mm *SRE* threshold is adopted here as a measure of ultimate benefit with the < 10 mm threshold a measure of utility. Table 4.5 shows the number and percentage of reconstructed points with an *SRE* falling below these thresholds. For the three laboratory dataset conditions $> 99.9\%$ of the reconstructed data was shown to fall within the lowest of these margins (< 5 mm). In the intraoperative dataset conditions, ***Intraoperative Case 1***, maintained similar performance of 92.3% of the reconstructed data falling within the < 5 mm threshold. ***Intraoperative Case 2*** and ***Intraoperative Case 3*** noticed a drop in performance, having 77.2% and 81.1% of the reconstructed data fall within the < 5 mm threshold respectively. When comparing to the lowest of bounds on the planned safety margins in laparoscopic surgeries, 10 mm, $\approx 100\%$ of the laboratory and $\approx 98\%$ of the intraoperative data lies within this tolerance. Further breakdown of the *SRE* distributions in the sub 5 mm range are shown in Table C.3 in Appendix C.

Dataset	Condition	Surface Registration Error (SRE) (mm) mean	std
Laboratory	Rotation-Only	1.0050	0.8237
	Translation-Only	1.1057	0.8873
	Translation+Rotation	1.0598	0.8376
Intraoperative	Case 1	1.9349	1.9486
	Case 2	3.2279	2.8418
	Case 3	3.2106	2.5355

Table 4.4: Mean and Standard Deviation of Surface Registration Error (SRE) on Datasets. Datasets that had a suitable *GO-ICP* registration candidate (all laboratory cases and ***Intraoperative Case 1***) are initialized using the *GO-ICP* registration candidate evaluated in Section 4.3.2. Otherwise, the manual registration candidate is used for initialization.

Dataset	Condition	Total Points	< 5 mm		< 10 mm		< 15 mm		< 20 mm		< 25 mm	
			# Points	%	# Points	%	# Points	%	# Points	%	# Points	%
Laboratory	Translation-Only	1354374	1353048	99.9	1354303	99.995	1354374	100	1354374	100	1354374	100
	Rotation-Only	1358820	1358250	99.96	1358820	100	1358820	100	1358820	100	1358820	100
	Translation+Rotation	1344523	1342778	99.9	1344523	100	1344523	100	1344523	100	1344523	100
Intraoperative	Case 1	735806	679505	92.3	730661	99.3	735713	99.99	735806	100	735806	100
	Case 2	2746115	2120084	77.2	2673421	97.4	2735364	99.6	2746100	99.9995	2746115	100
	Case 3	2732357	2216546	81.1	2673512	97.8	2724510	99.7	2731565	99.97	2732357	100

Table 4.5: Surface Registration Error (SRE) - Data Within Surgical Tolerance. The total number of reconstructed points as well as the number and % of reconstructed points falling within the $< 5 - 25$ mm *SRE* thresholds are reported here. Datasets that had a suitable *GO-ICP* registration candidate (all laboratory cases and ***Intraoperative Case 1***) are initialized using the *GO-ICP* registration candidate evaluated in Section 4.3.2. Otherwise, the manual registration candidate is used for initialization.

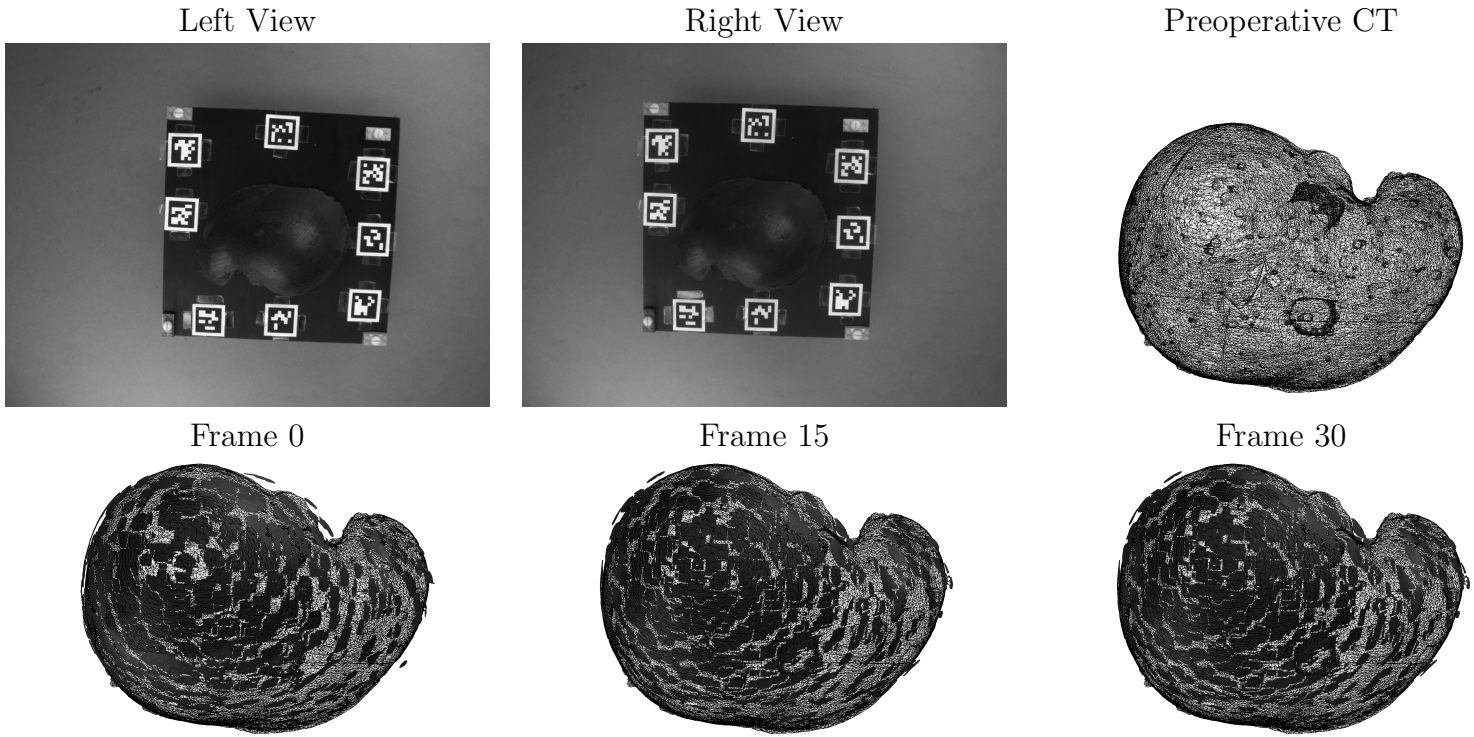


Figure 4.17: Qualitative Surface Registration Error (SRE) Results - Translation-Only. The left and right (non-rectified) views from the stereo camera are shown at the initial frame (frame 0). The preoperative *CT* scan is shown in a top-down view. The recovered stereo reconstructions (shown in darkest grey), registered to the preoperative *CT* scan, are shown at frames 0, 15, and 30.

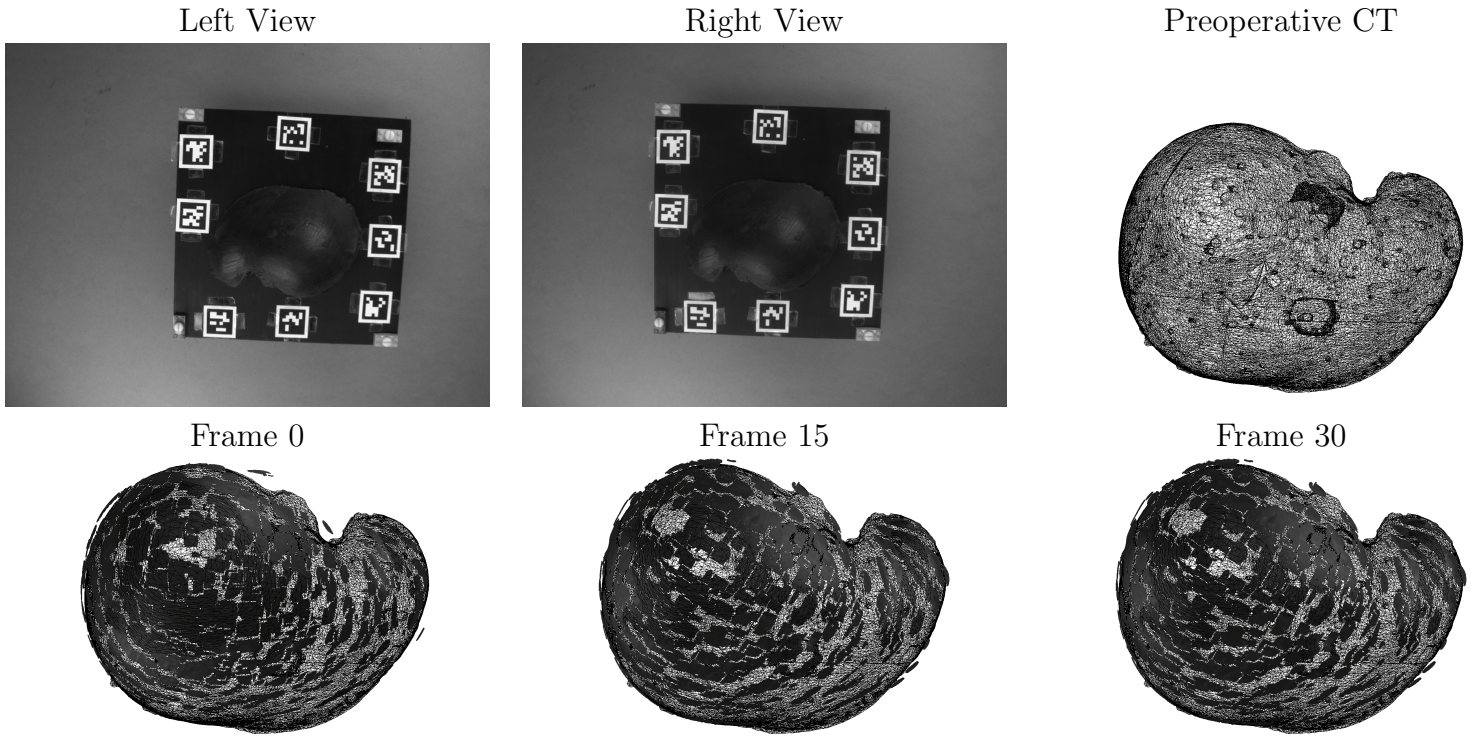


Figure 4.18: Qualitative Surface Registration Error (SRE) Results - Rotation-Only. The left and right (non-rectified) views from the stereo camera are shown at the initial frame (frame 0). The preoperative CT scan is shown in a top-down view. The recovered stereo reconstructions (shown in darkest grey), registered to the preoperative CT scan, are shown at frames 0, 15, and 30.

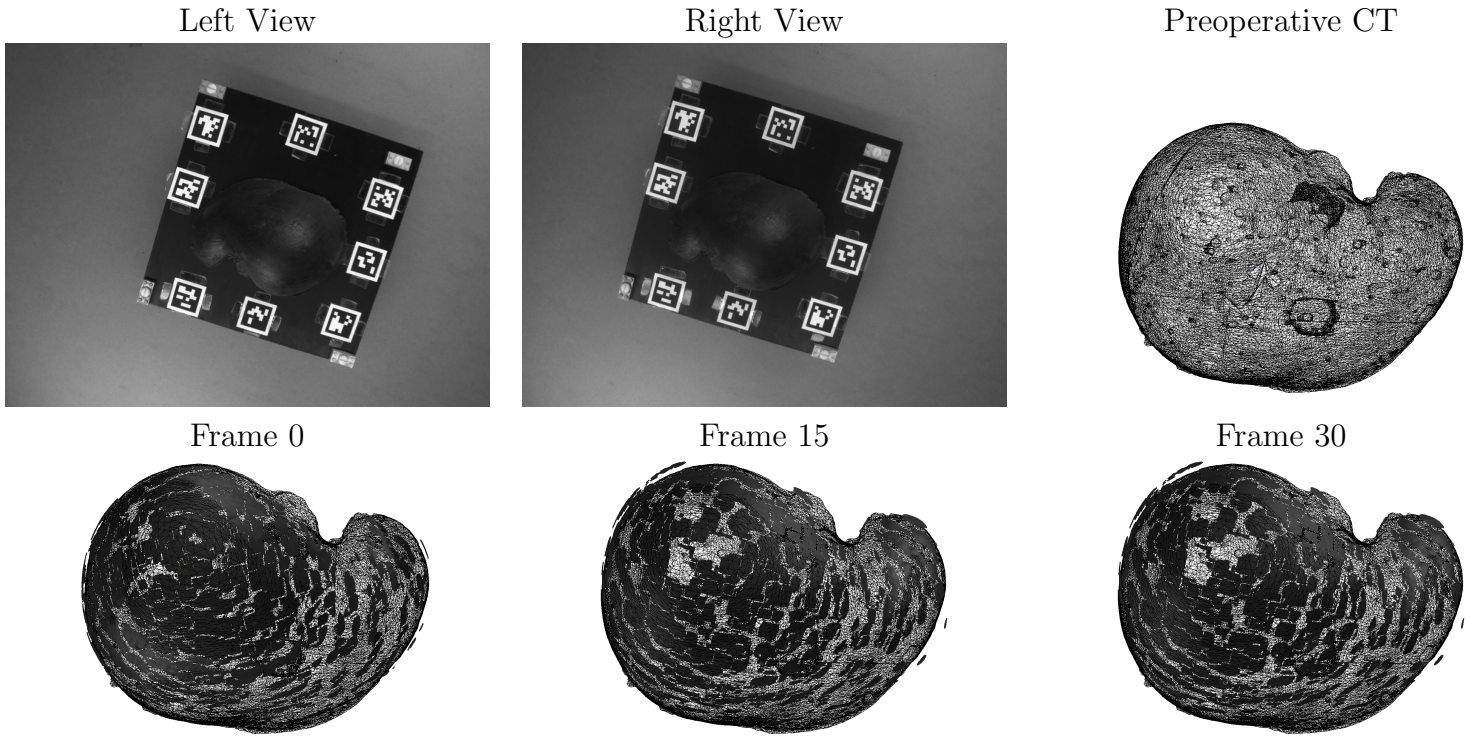


Figure 4.19: Qualitative Surface Registration Error (SRE) Results - Translation+Rotation. The left and right (non-rectified) views from the stereo camera are shown at the initial frame (frame 0). The preoperative *CT* scan is shown in a top-down view. The recovered stereo reconstructions (shown in darkest grey), registered to the preoperative *CT* scan, are shown at frames 0, 15, and 30.

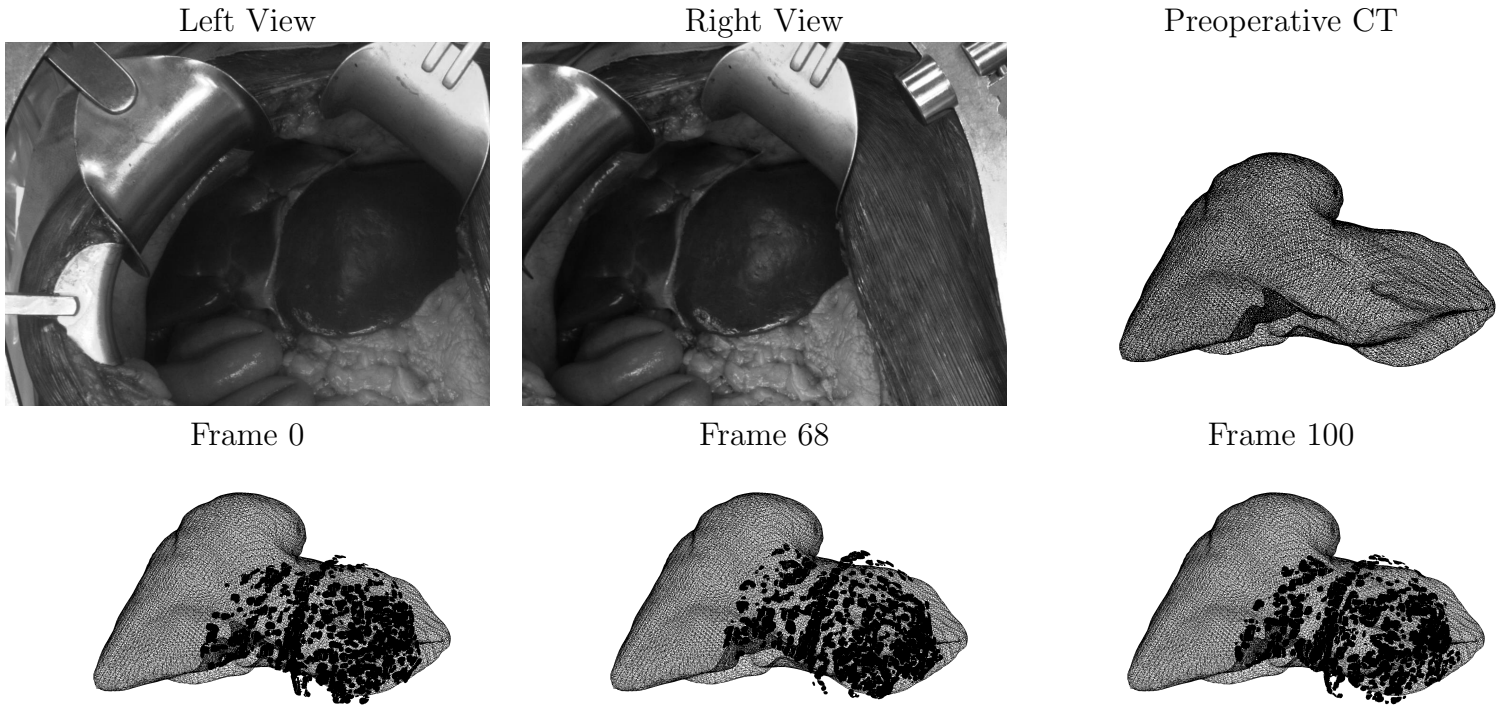


Figure 4.20: Qualitative Surface Registration Error (SRE) Results - Intraoperative Case 1. The left and right (non-rectified) views from the stereo camera are shown at the initial frame (frame 0). The preoperative CT scan is shown with the anterior face of the liver facing the camera. The recovered stereo reconstructions (shown in darkest grey), registered to the preoperative CT scan, are shown at frames 0, 68, and 100.

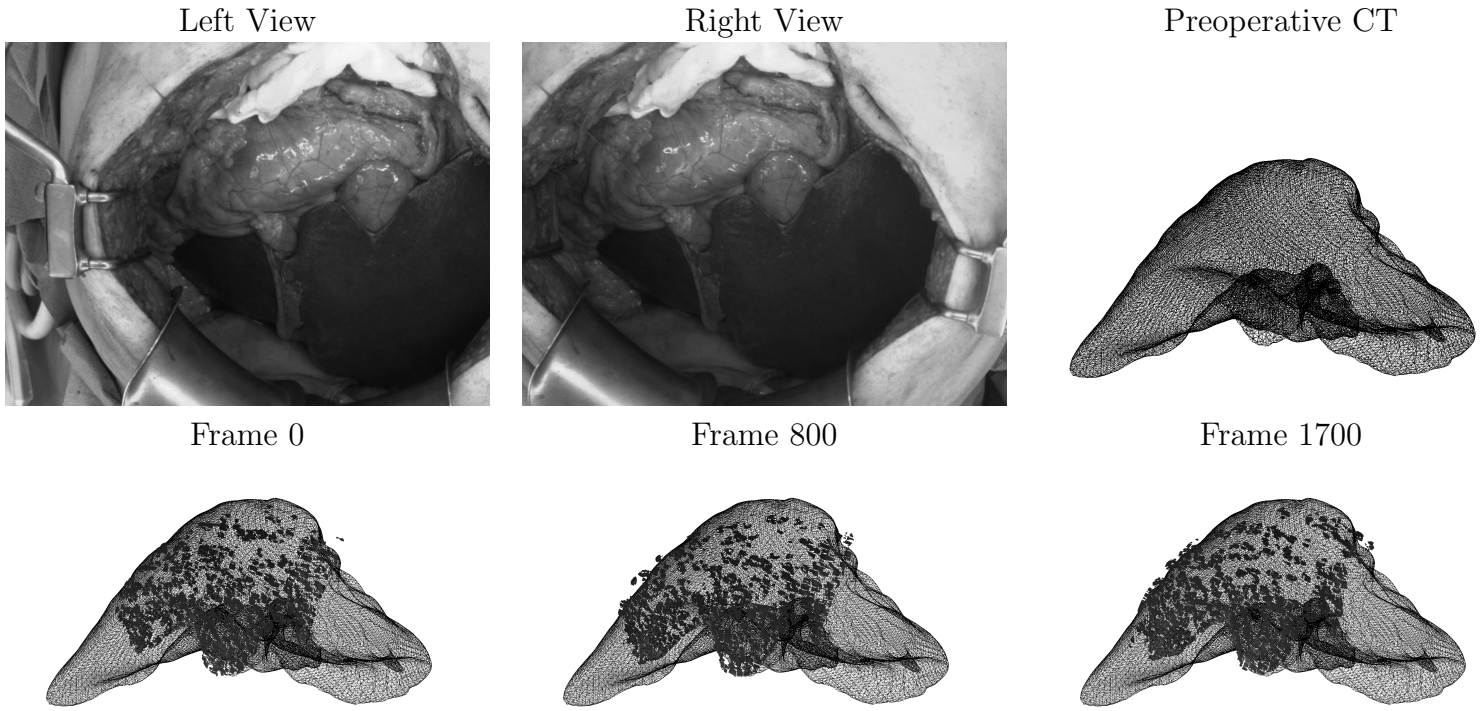


Figure 4.21: Qualitative Surface Registration Error (SRE) Results - Intraoperative Case 2. The left and right (non-rectified) views from the stereo camera are shown at the initial frame (frame 0). The preoperative CT scan is shown with the anterior face of the liver facing the camera. The recovered stereo reconstructions (shown in darkest grey), registered to the preoperative CT scan, are shown at frames 0, 800, and 1700.

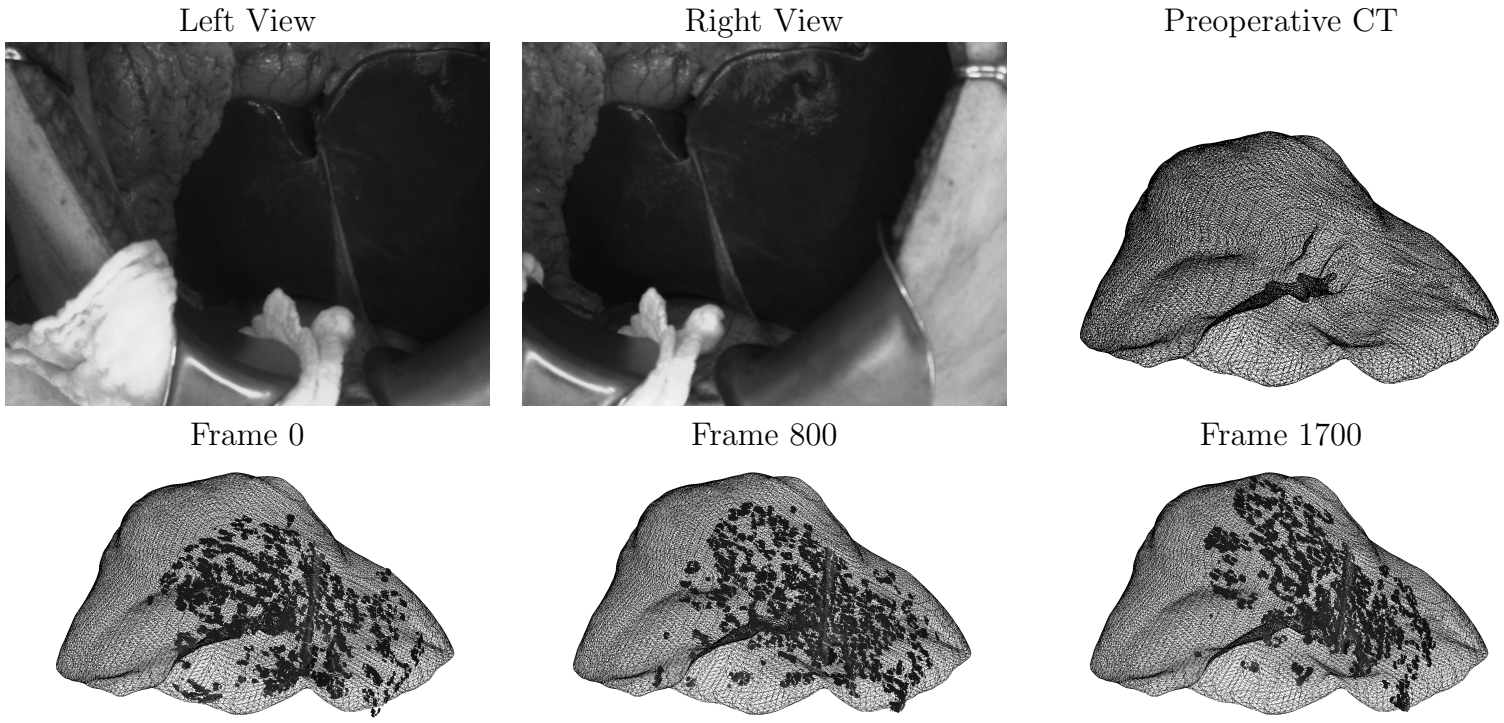


Figure 4.22: Qualitative Surface Registration Error (SRE) Results - Intraoperative Case 3. The left and right (non-rectified) views from the stereo camera are shown at the initial frame (frame 0). The preoperative CT scan is shown with the anterior face of the liver facing the camera. The recovered stereo reconstructions (shown in darkest grey), registered to the preoperative CT scan, are shown at frames 0, 800, and 1700.

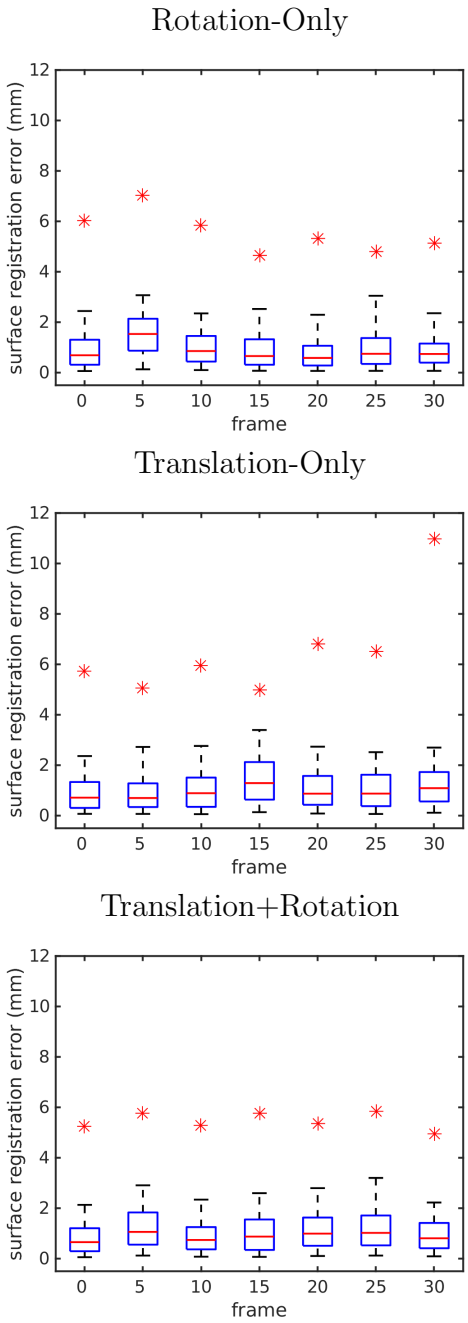


Figure 4.23: Surface Registration Error (SRE) - Laboratory Dataset. *SRE*, Equation (4.1), is shown at five degree/mm increments across the laboratory sequences. Box plots are shown for each of the reported frames. Blue boxes cover the interquartile range, with the red line showing the median value. Whiskers cover 90% of the reconstructed point clouds. The largest outlier for each dataset is indicated by a red asterisk.

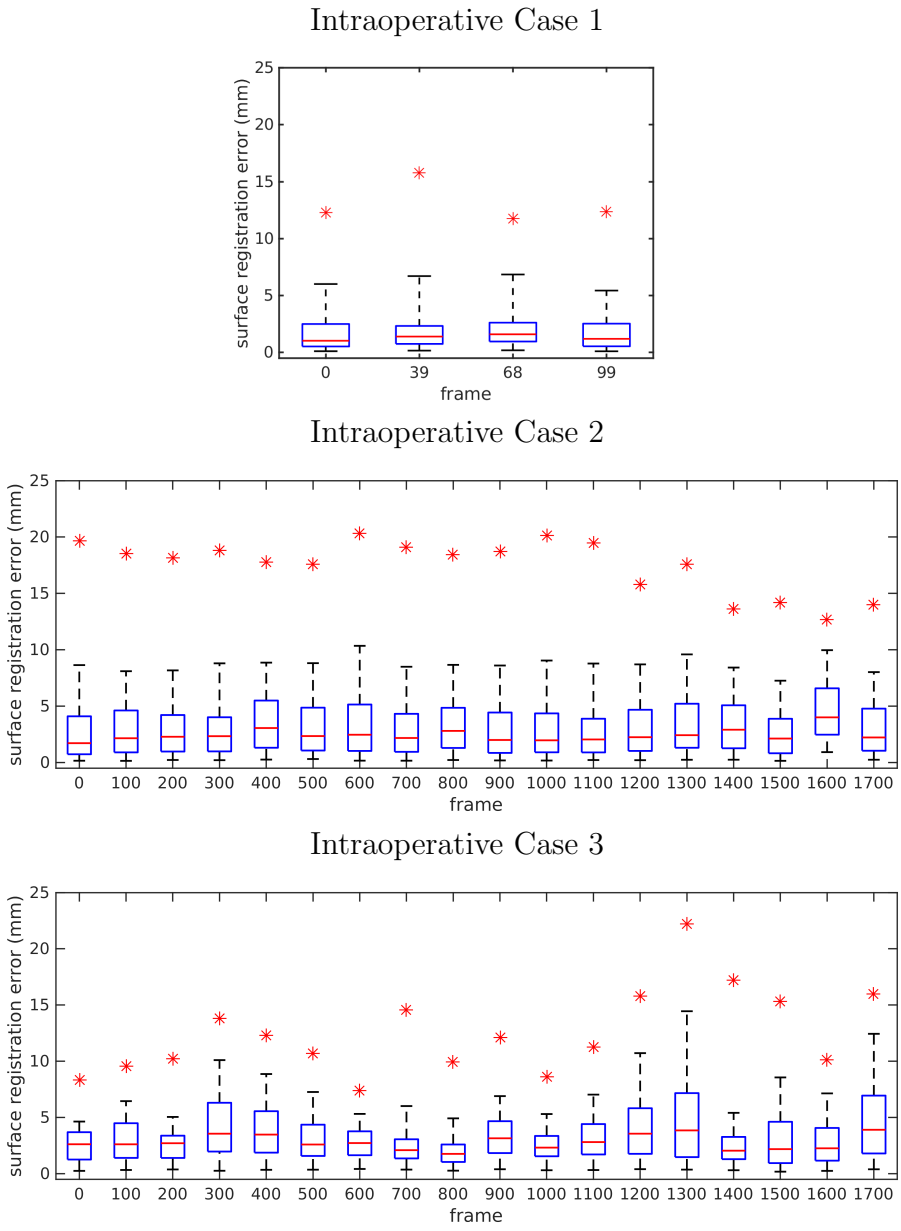


Figure 4.24: Surface Registration Error (SRE) - Intraoperative Dataset. *SRE*, Equation (4.1), is shown in ***Intraoperative Case 1*** for four frames representing various amounts of respiratory motion that were chosen to show the system performance over the largest portion of the organ's motion. *SRE* is shown at 100 frame increments across the longer intraoperative sequences: ***Intraoperative Case 2*** and ***Intraoperative Case 3***. Box plots are shown for each of the reported frames. Blue boxes cover the interquartile range, with the red line showing the median value. Whiskers cover 90% of the reconstructed point clouds. The largest outlier for each dataset is indicated by a red asterisk.

4.6 Discussion

In this chapter we provided an evaluation of the proposed *SN* system and some of its components in isolation. The system was evaluated both in controlled laboratory conditions, where a liver phantom underwent a series of controlled motion patterns as well as in a number of open liver resection tasks.

4.6.1 Initialization

Two initialization methods were compared: (i) a manual initialization procedure and (ii) a semi-automated registration leveraging *GO-ICP*. The modified *GO-ICP* approach was able to provide a meaningful "one-shot" registration candidate in the laboratory dataset conditions and a meaningful candidate, as the 14th candidate, in one of the three intraoperative dataset conditions. As the *GO-ICP* algorithm attempts to find a rigid registration between the source and the target point cloud, it is intuitive that the ability to find a successful registration candidate will depend in part on the amount of deformation present between the preoperative state of the organ and its intraoperative state. With the level of deformation being negligible (if any) in the laboratory dataset test conditions a delineation between the performance of *GO-ICP* on the laboratory dataset and the intraoperative dataset can be expected.

Another hurdle present in the use of *GO-ICP* in this application domain is the large degree of local shape similarity. By its very nature, the algorithm cannot distinguish between perfect symmetries [208]. When looking at the volume of a typical liver as a whole, self-symmetry is unlikely. However, when analyzing the particular use case of *SN*, intraoperatively recovered estimates of the organ surface

typically only cover a small portion of the surface of the organ. The recovered surface is typically smooth and without any uniquely identifiable landmarks to guide 3D registration. As such, this problem is also susceptible to issues of symmetry. ***Intraoperative Case 1*** illustrates this problem, despite being the single intraoperative case having some success with the proposed semi-automated initialization. Upon visual inspection, as seen in Figures 4.11 and C.3, there are a great number of locations providing a viable registration between the reconstruction and the preoperative *CT* when based solely on the shape of the two objects. For example, under the conditions provided to the *GO-ICP* search, registration candidates 1-13, while not physically meaningful, each had a lower *SRE* than the visibly plausible registration location provided by registration candidate 14. In cases with minimal deformation present, increased surface coverage is perhaps the only solution for fully breaking such issues of symmetry.

To our knowledge there is only one previous attempt to automate this portion of the typical *SN* pipeline [149]. This previous approach relies on 3D feature descriptors that could be localized in both the preoperative imagery as well as in the intraoperative surface reconstruction. It also relies on pre-segmenting the *CT* scan to only contain the visible portion of the surface and manual delineation of the liver contour in pre- and intraoperative imagery. While impossible to make direct comparison, our proposed method achieved a mean *SRE* of 0.82-0.91 mm across the laboratory dataset conditions. The most similar comparison point would be to the tests performed in the previous work where they took 43% of the *CT* surface (roughly in the same location as our stereo reconstruction), applied no deformation to the sliced points, applied a random transformation and attempted to register the partial surface. This resulted in an approximate error of 10 mm

(without the added difficulty of having noise from the stereo reconstruction of the surface, which was shown to dramatically increase error in other tests the authors performed). Our experiments also seem to provide the first quantitative results for semi-automated initialization on intraoperative data (showing a mean *SRE* of 1.79 mm on ***Intraoperative Case 1***). Further, the *SRE* for the semi-automated approach performed favourably compared to the fully manual approach, which yielded an *SRE* of 1.65 mm; see Appendix C, Table C.1. Similarly, comparable *SRE* results were found for semi-automated vs. fully manual initializations on the laboratory datasets (Table 4.4 vs. C.1). Overall, these results suggest promise for our semi-automated initialization procedure, at least for cases where superior coverage in surface reconstruction is achieved.

4.6.2 Motion estimation

The first component of the end-to-end system evaluation looked at evaluation of the extracted motion of the organ in comparison to a known motion of the organ in a series of controlled laboratory evaluations. As mentioned previously, the largest component of motion in open liver surgery conditions is due to respiration causing large translational motion in the order of $\approx 10-75$ mm [33]. The laboratory dataset covers average translational motions, in the range of $0-30$ mm, and greater than average rotational motions, in the range of $0-30$ degrees. To our knowledge it is the first liver motion evaluation (simulated, laboratory or in/ex-vivo) in an open surgical setting. We also believe it to be the first focused on the motion of the liver as opposed to the relative camera motion (which has been seen as a stream of evaluation in several *MIS* applications). Over our three laboratory dataset

conditions, the recovery of the translational component and rotational component of the organ motion was within ranges reported previously [130, 127, 123, 56]. While the results reported elsewhere typically derive from *MIS* procedures, and therefore from somewhat different conditions than we target, they nevertheless provide the closest points for comparison.

4.6.3 3D surface estimation and optical organ registration

Surface Registration Error (SRE) has been reported to evaluate the quality of the registration between the recovered surface models and volumetric models. For each dataset, mean *SRE* was in the range $1.03 - 3.23$ mm (Table 4.4). It also was seen that 99.9% and 80.7% of the registered laboratory and intraoperative points (respectively) yielded under 5 mm error, while $\approx 100\%$ and 97.8% were under 10 mm error (Table 4.5). Further, time series of *SRE* distributions showed no tendency toward drift (Figures 4.23 and 4.24), which provides evidence that the system can maintain registration across time.

SRE and surgical tolerances have an important relationship for *SN* systems seeking adoption in the *Operating Room (OR)*. As surgical tolerances decrease, so too must the registration error of systems operating in these scenarios. With surgical tolerances in laparoscopic interventions documented in the $10 - 25$ mm range [140], it is clear that registration errors must fall within this range to be part of a deployed *SN* solution. We also noted that a more restrictive threshold of < 5 mm has been adopted in several publications in the *SN* literature. This threshold has been reasoned from several different positions including relation to accepted tolerances in other portions of the anatomy [93], ability to restore spatial

understanding to the surgeon [158], and in comparison to the limits in place around typical surgical margins [14]. Our reported errors lie within these tolerances.

Notably, the only two *Food and Drug Administration (FDA)*-approved liver surgery systems, CAS-One[5] [137] and Explorer[6] [87], do not compensate for real-time motion [175]. If disturbed, either through patient respiration or surgical manipulation, accuracy is negatively impacted and re-initialization is needed. In practice, current systems are typically used at breath hold. As such, real-time localization and feedback is needed in an ideal system [175]. More generally, commercial systems rely on optical tracking, which has a stated accuracy of < 2 mm, as evaluated at a single point in time [131]. The errors reported are within comparable bounds and the presented system is capable of providing continual updates. Thus, our system has potential to provide the precise anatomical location of tumours in the complex vasculature, in real-time, as the liver undergoes motion throughout the course of surgery.

[5]CAS-One was developed at the University of Bern and is manufactured by CAScination, Bern, Switzerland.

[6]Explorer is manufactured by Pathfinder Therapeutics, Inc., Nashville, TN, United States.

Chapter 5

Conclusions

5.1 Summary

An end-to-end system for fast and accurate 3D surface reconstruction and motion estimation for alignment with a preoperative volumetric scan has been presented. The system has been designed and evaluated for application to liver surgery in an open setting, with open surgery still the dominant setting for liver operations [166]. It is comprised of three key components: initialization, 3D surface recovery, and 3D motion estimation. The system has been evaluated empirically with reference to novel laboratory and intraoperative datasets, with results showing that performance is within tolerances expected for integration into *SN* systems.

The first system component, initial registration of preoperative *CT* information and intraoperative surface reconstructions, was provided using two methods: manual initialization using a *GUI* as well as a semi-automated initialization strategy based on a modified version of *GO-ICP* [208]. Manual initialization strategies are common practice in the *SN* community. Despite being a reliable method for obtaining

reasonable initializations for most cases, manual initialization strategies have been shown to interrupt surgical workflow and the accuracy of the resulting registration is highly dependent on the expertise of the user performing the initial registration [14]. The pitfalls inherent in the manual initialization process motivate the creation of new strategies that improve this aspect of *SN* systems. The proposed semi-automated initialization strategy has potential to improve workflow through a decreased reliance on expensive user input to system operation.

As its second major component, the system integrates an adaptive *CTF* stereo algorithm [179] capable of fast and accurate 3D surface reconstruction of the surface of the visible anatomy. The approach yields data-driven dense reconstruction in a computationally-efficient manner by allowing coarse resolution image data to inform fine resolution reconstruction, even in low texture regions [179]. Typical applications of *CTF* processing in stereo imagery suffers from blurring about 3D discontinuities [179]. *CTF* processing has been suggested before for use in *SN* applications [189]. Complementary use of adaptive windows in our approach allows for precise reconstruction of 3D boundaries [179]. With important information for downstream processing occurring at these discontinuities (e.g., organ contours for pre- to intraoperative registration) as well as the downstream benefits of employing strategies to maximize surface coverage, this advance is an important distinction that we believe is a first in the *SN* community.

The third major system component makes use of a robust, 3D motion estimator based on interframe feature matching to register a time series of reconstructions back to the initial frame for subsequent registration back to a volumetric *CT* scan. This approach proved to be stable, with no observable motion drift in the reported motion errors or systematic increase of *SRE* in full sequence registration evaluations.

When compared to other systems that may chain a series of reconstructions to a single keyframe [193, 129, 11], which rapidly lead to drift, or several of the *SLAM*-based systems where localization ability deteriorates as they go further from their starting position [130, 37, 56], this approach shows great promise for scenarios where it is applicable (such as the designed scenario of intraoperative planning). Also, unlike typical approaches in the *SN* community that use *ICP* for incremental 3D model registration [193, 185, 34, 11, 113], the proposed feature matching-based approach encourages registration to bring physically meaningful features into alignment. Typical *ICP*-based registration minimizes the L_2-error between the two point sets and is prone to sliding effects, where the optimal registration incrementally drifts away from the originally hand-picked/physically meaningful location while the two surfaces (typically smooth in nature) slide along each other due to subtle changes in the error profiles.

Beyond the system per se, this research also addresses a pressing need in the *SN* community, a controlled laboratory dataset allowing for ground-truth evaluation of the motion estimation ability of *SN* systems was created. To our knowledge this is the first controlled dataset placed in open surgical analogue settings with ground-truth signals for the motion of the phantom liver. This addresses two needs of the *SN* community. First, it provides a controlled manner with which to test *SN* systems in an open surgical setting. As mentioned previously, the vast majority of liver resection tasks are performed in an open surgical setting. Second, the laboratory dataset has been shared with the *SN* community addressing yet another need by allowing for future comparisons between proposed systems to become a possibility. The utility of the dataset has been demonstrated via evaluation of the developed system. Here, the results suggest the potential for integration into a

clinical system.

5.2 Notable limitations

While the *SN* system presented in this dissertation is capable of providing accurate results in an efficient manner, the system does have its limitations. There are three notable limitations of the system design. First, the system does not currently account for any sources of non-rigid deformation of the liver. Interframe motion estimation currently makes use of previous observations, namely that the liver behaves in a predominantly rigid fashion during intraoperative inspection [173]. While these assumptions provide compelling results under the evaluated scenarios, extending the use of the system into further phases of the operation is likely to result in these assumptions being violated because of non-rigid deformations taking place due to interactions with the liver by the surgical staff. Furthermore, there is often a non-rigid deformation of the organ that occurs between when the preoperative image is acquired and when the intraoperative stereo imagery is initially obtained. This later source of deformation likely accounts for a noticeable portion of the reported *SRE* in the evaluations provided in this dissertation. Providing a physically meaningful deformation model that is able to run within the constraints of a typical surgical workflow should be considered in future iterations of this work.

Second, the semi-automated initialization strategy proposed in this dissertation worked extremely well in laboratory conditions but struggled to find a suitable registration candidate in two of the intraoperative datasets. The currently implemented semi-automated initialization strategy makes use of an underlying technique that computes the n-best registration candidates, in terms of *SRE*. As observed in the

results, the "best" registration candidate in terms of physical meaningfulness may not align with the best potential alignment in terms of *SRE* (due to issues such as measurement error, limited surface coverage combined with a large degree of local shape similarity, and deformation between the intraoperative and preoperative states of the liver). Determining additional constraints and/or methodologies that may help the semi-automated initialization technique find a physically meaningful registration in a suitable amount of time with minimal effort from the user should be explored in future work.

Third, although not experienced in the data obtained for evaluating the system, interacting with the organ of interest during the running of the system has the potential to induce feature drop off that could ultimately cause interframe feature tracking to fail. Extending the interframe tracking module to be able to recover from tracking failures and/or decrease the potential for long-term drift should be explored in future iterations of this work.

In addition, there were two notable limitations in the evaluation of the *SN* system presented in this work. First, error analysis is provided only at the surface when structures of interest for this application are often embedded within the organ volume. Integrating subsurface imaging into the intraoperative data being evaluated could provide additional information about subsurface accuracy of the system and should be considered as a part of future work.

Second, additional intraoperative information capturing increased variation in patient anatomy, surgical workflow and in the extent, location and type of disease as well as capturing video sequences over extended durations during the surgical intervention would be beneficial for further evaluation of system performance. Furthermore, data driven design principles could be used to help focus development

efforts by identifying existing issues present in the system design that were not uncovered by the current intraoperative sequences as well as new use cases the system will need to deal with in deployment.

5.3 Future directions

5.3.1 Support for non-rigidity

To date, introduction of non-rigid registration or refinement has not been explored in the context of the proposed system. However, three places exist in the presented framework to support research into non-rigid registration of an intraoperative 3D reconstruction and preoperative volumetric imagery. First, modification of the *GO-ICP BnB* constraints to include non-rigidity could support recovery of corresponding relationships between the initial frame and preoperative *CT* scan. Second, a bootstrap approach of providing an additional non-rigid refinement as a postprocessing step could provide this support. Third, fully utilizing the 3D deformation field (Section 3.4) to support non-rigid registrations of sequential 3D surface reconstructions could provide end-to-end support for non-rigid alignment of data. Note that both the rigid and non-rigid components of the motion of the surface are captured by the deformation fields currently maintained in the system.

5.3.2 Improve the semi-automated initialization strategy

Currently the *GO-ICP* algorithm at the core of the semi-automated initialization uses the L_2-error present between the closest points in the two point sets as a minimization criteria. In the presence of non-rigidity this choice can become

problematic. A natural extension would be to relax the L_2-error currently supported by the algorithm to allow for some degree of non-rigidity as done elsewhere [149]. Furthermore, introduction of additional constraints placed by the liver contour [37, 38, 149] could be a powerful means of resolving some of the issues currently observed with partial observation of the surface and the large degree of local shape similarity present at those levels of surface coverage. The warped region of interest mask already maintained by the system, Section 3.3.3, could be used to provide this organ boundary/occlusion information.

Recently, an approach to global rigid registration that uses a quasi-*BnB* strategy has been proposed in the literature [44]. The quasi-*BnB* replaces the linear lower bounds seen in *GO-ICP* with quadratic quasi-lower bounds providing a significant speed up (especially when high registration accuracy is required). Integration of quasi-*BnB* constraints into the automated registration strategy could allow for a further decrease in disruption to surgical workflow and/or increased fidelity in the registration result in the same computational budget.

5.3.3 Feature track recovery for long-term tracking and occlusion tolerance

The interframe motion estimation module in the *SN* system makes use of 2D feature tracking methods. The use of 2D feature tracks, when combined with the computed disparity information at each frame, allows for a registration of any frame back to the initial frame in the image sequence. This design provides several benefits in comparison to many competing methods; however, it is not without its disadvantages. A known downfall of 2D feature tracking methods is that of feature

drop off as tracking progresses throughout the video sequence. Drop off can occur for a variety of reasons including occlusion of the feature in a given frame, large motions that are outside of the capture window and/or induce significant motion blur, and any significant change in appearance of the feature location. Making use of feature matching techniques to help re-establish lost feature tracks should be explored as a form of tracking recovery. Furthermore, extensions which exploit the periodicity observed in the data due to the patient's respiration may also be of interest to future work.

5.3.4 Subsurface imaging and error analysis

An extension of the presented approach to include subsurface registration through the use of additional sensing modalities that support subsurface data acquisition during surgery (e.g., *Ultrasound (US)*, *Cone Beam Computed Tomography (CBCT)*, and open *Magnetic Resonance Imaging (MRI)*) is also of interest. Appropriate application of intraoperative subsurface scanning and registration constraints may provide better localization for surgeons when surgical margins are tight. Furthermore, integrating subsurface information into the surface-based registration approach is an important step for being able to measure the correlation between measurements of registration errors at the surface of the organ and the ability to target subsurface structures accurately.

5.3.5 Additional intraoperative evaluation

Additional testing on clinical data is desired to further validate and develop the presented techniques. If possible, the ability to have external validation in the

operating room conditions could yield further information regarding the accuracy of the proposed system. Introduction of fiducials for a controlled set of intraoperative cases could provide measures of *Target Registration Error (TRE)* that would further allow for the evaluation of the initialization and motion estimation portions of the system. Data collected in conditions capturing extra surface coverage of the liver could provide a starting point for aspects of non-rigid refinement. Here, surface coverage is often a necessary (but not sufficient) condition for obtaining any meaningful deformation corrections in these settings [173, 158].

Testing within the *OR*, in real-time, will be another area of interest for future testing. To date, the use of the *SN* system inside the *OR* has not been possible due to need for further ethics and regulatory approvals. However, use of the system within the surgical theatre will be an important step towards gathering feedback from surgeons on the utility of this system and determining further avenues for system improvement. The ultimate goal would be to put the system through clinical trials in order to quantify the impact of the system on surgical outcomes.

Bibliography

[1] M. Alexa, J. Behr, D. Cohen-Or, S. Fleishman, D. Levin, and C. T. Silva, "Computing and rendering point set surfaces," *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 9, no. 1, pp. 3–15, Jan. 2003. 80

[2] H. O. Altamar, R. E. Ong, C. L. Glisson, D. P. Viprakasit, M. I. Miga, S. D. Herrell, and R. L. Galloway, "Kidney Deformation and Intraprocedural Registration: A Study of Elements of Image-Guided Kidney Surgery," *Journal of Endourology*, vol. 25, no. 3, pp. 511–517, Mar. 2011. 48

[3] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-Squares Fitting of Two 3-D Point Sets," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 9, no. 5, pp. 698–700, Sep. 1987. 87, 93

[4] N. Atzpadin, P. Kauff, and O. Schreer, "Stereo Analysis by Hybrid Recursive Matching for Real-Time Immersive Video Conferencing," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 14, no. 3, pp. 321–334, Mar. 2004. 29, 30

[5] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "VoxelMorph: a learning framework for deformable medical image registration," *IEEE Transactions on Medical Imaging (TMI)*, vol. 38, no. 8, pp. 1788–1800, Aug. 2019. 38

[6] J. Bano, S. A. Nicolau, A. Hostettler, C. Doignon, J. Marescaux, and L. Soler, "Registration of Preoperative Liver Model for Laparoscopic Surgery from Intraoperative 3D Acquisition," in *Augmented Reality Environments for Medical Imaging and Computer-Assisted Interventions (AE-CAI)*. Nagoya, Japan: Springer, Berlin, Heidelberg, 2013, pp. 201–210. 16, 48

[7] M. Baumhauer, T. Simpfendörfer, B. P. Muller-Stich, D. Teber, C. N. Gutt, J. Rassweiler, H. P. Meinzer, and I. Wolf, "Soft tissue navigation for laparoscopic partial nephrectomy," *International Journal of Computer Assisted Radiology and Surgery (IJCARS)*, vol. 3, no. 3-4, pp. 307–314, May 2008. 17, 18, 43, 46, 47, 73

[8] H. Bay, T. Tuytelaars, and L. van Gool, "SURF: Speeded Up Robust Features," in *European Conference on Computer Vision (ECCV)*, Graz, Austria, 2006, pp. 404–417. 41, 83

[9] J.-C. Bazin, Y. Seo, and M. Pollefeys, "Globally Optimal Consensus Set Maximization Through Rotation Search," in *Asian Conference of Computer Vision (ACCV)*, Daejeon, Korea, 2012, pp. 539–551. 65

[10] S. Belongie, J. Malik, and J. Puzicha, "Shape Matching and Object Recognition Using Shape Contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 24, no. 24, pp. 509–522, Apr. 2002. 64

[11] C. Bergeles, P. Pratt, R. Merrifield, A. Darzi, and G.-Z. Yang, "Multi-view Stereo and Advanced Navigation for Transanal Endoscopic Microsurgery," in *Medical Image Computing and Computer Assisted Interventions (MICCAI)*, Cambridge, MA, USA, 2014, pp. 332–339. 49, 76, 135

[12] S. Bernhardt, J. Abi-Nahed, and R. Abugharbieh, "Robust Dense Endoscopic Stereo Reconstruction for Minimally Invasive Surgery," in *Medical Image Computing and Computer Assisted Interventions (MICCAI)*, Nagoya, Japan, 2013, pp. 254–262. 27, 76

[13] S. Bernhardt, S. p. A. Nicolau, V. Agnus, L. Soler, C. Doignon, and J. Marescaux, "Automatic localization of endoscope in intraoperative CT image: A simple approach to augmented reality guidance in laparoscopic surgery," *Medical Image Analysis (MIA)*, vol. 30, pp. 130–143, May 2016. 33, 46

[14] S. Bernhardt, S. A. Nicolau, L. Soler, and C. Doignon, "The Status of Augmented Reality in Laparoscopic Surgery as of 2016," *Medical Image Analysis (MIA)*, vol. 37, pp. 66–90, Apr. 2017. 3, 8, 9, 33, 46, 47, 48, 52, 57, 117, 132, 134, 193

[15] P. J. Besl and H. D. McKay, "A Method for Registration of 3-D Shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 14, no. 2, pp. 239–256, 1992. 30, 42, 44, 62, 63, 180, 192

[16] R. Bhalodia, S. Y. Elhabian, L. Kavan, and R. T. Whitaker, "A Cooperative Autoencoder for Population-Based Regularization of CNN Image Registra-

tion," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Shenzhen, China, 2019, pp. 391–400. 38

[17] S. Birchfield and C. Tomasi, "A pixel dissimilarity measure that is insensitive to image sampling," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 20, no. 4, pp. 401–406, Apr. 1998. 40

[18] M. J. Black and P. Anandan, "The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields," *Computer Vision and Image Understanding (CVIU)*, vol. 63, no. 1, pp. 75–104, Jan. 1996. 80

[19] G. Blais and M. D. Levine, "Registering Multiview Range Data to Create 3D Computer Objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 17, no. 8, pp. 820–824, Aug. 1995. 64

[20] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000. 31

[21] P. Brandao, E. Mazomenos, A. Rau, M. Janatka, and D. Stoyanov, "Widening siamese neural networks for stereo matching in colonoscopy," in *Joint Workshop on New Technologies for Computer/Robot Assisted Surgery (CRAS)*, Montpellier, France, 2017. 38

[22] K. K. Brock, M. Hawkins, C. Eccles, J. L. Moseley, D. J. Moseley, D. A. Jaffray, and L. A. Dawson, "Improving Image-Guided Target Localization Through Deformable Registration," *Acta Oncologica*, vol. 47, no. 7, pp. 1279–1285, Jul. 2009. 48

[23] M. Brown and D. G. Lowe, "Invariant Features from Interest Point Groups." in *British Machine Vision Conference (BMVC)*, Cardiff, UK, 2002, pp. 656–665. 49

[24] N. C. Buchs, F. Volonte, F. Pugin, C. Toso, M. Fusaglia, K. Gavaghan, P. E. Majno, M. Peterhans, S. Weber, and P. Morel, "Augmented environments for the targeting of hepatic lesions during image-guided robotic liver surgery," *Journal of Surgical Research*, vol. 184, no. 2, pp. 825–831, Oct. 2013. 4

[25] D. Burschka, M. Li, M. Ishii, R. H. Taylor, and G. D. Hager, "Scale-Invariant Registration of Monocular Endoscopic Images to CT-scans for Sinus Surgery," *Medical Image Analysis (MIA)*, vol. 9, no. 5, pp. 413–426, Jul. 2005. 3, 43

[26] D. M. Cash, M. I. Miga, S. C. Glasgow, B. M. Dawant, L. W. Clements, Z. Cao, R. L. Galloway, and W. C. Chapman, "Concepts and Preliminary Data Toward the Realization of Image-guided Liver Surgery," *Journal of Gastrointestinal Surgery (JOGS)*, vol. 11, no. 7, pp. 844–859, Apr. 2007. 51

[27] D. M. Cash, M. I. Miga, T. K. Sinha, R. L. Galloway, and W. C. Chapman, "Compensating for intraoperative soft-tissue deformations using incomplete surface data and finite elements," *IEEE Transactions on Medical Imaging (TMI)*, vol. 24, no. 11, pp. 1479–1491, 2005. 15

[28] P.-L. Chang, D. Stoyanov, A. J. Davison, and P. E. Edwards, "Real-Time Dense Stereo Reconstruction Using Convex Optimisation with a Cost-Volume for Image-Guided Robotic Surgery," in *Medical Image Computing and Computer Assisted Interventions (MICCAI)*, Nagoya, Japan, 2013, pp. 42–49. 27, 76

[29] Y. Chen and G. Medioni, "Object Modelling by Registration of Multiple Range Images," *Image and Vision Computing (IVC)*, vol. 10, no. 3, pp. 145–155, 1992. 30, 44, 62, 63, 180, 192

[30] T. Ching, D. S. Himmelstein, B. K. Beaulieu-Jones, A. A. Kalinin, B. T. Do, G. P. Way, E. Ferrero, P.-M. Agapow, M. Zietz, M. M. Hoffman, W. Xie, G. L. Rosen, B. J. Lengerich, J. Israeli, J. Lanchantin, S. Woloszynek, A. E. Carpenter, A. Shrikumar, J. Xu, E. M. Cofer, C. A. Lavender, S. C. Turaga, A. M. Alexandari, Z. Lu, D. J. Harris, D. DeCaprio, Y. Qi, A. Kundaje, Y. Peng, L. K. Wiley, M. H. S. Segler, S. M. Boca, S. J. Swamidass, A. Huang, A. Gitter, and C. S. Greene, "Opportunities and obstacles for deep learning in biology and medicine," *Journal of The Royal Society Interface*, vol. 15, no. 141, p. 20170387, 2018. 38

[31] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, and G. Ranzuglia, "MeshLab: an Open-Source Mesh Processing Tool," in *Sixth Eurographics Italian Chapter Conference*, Salerno, Italy, 2008, pp. 129–136. 62

[32] L. W. Clements, W. C. Chapman, B. M. Dawant, R. L. Galloway, Jr., and M. I. Miga, "Robust surface registration using salient anatomical features for image-guided liver surgery: Algorithm and validation," *Medical Physics (MP)*, vol. 35, no. 6, pp. 2528–2540, May 2008. 51

[33] L. W. Clements, P. Dumpuri, W. C. Chapman, B. M. Dawant, R. L. Galloway, and M. I. Miga, "Organ Surface Deformation Measurement and Analysis in Open Hepatic Surgery: Method and Preliminary Results From 12 Clinical

Cases," *IEEE Transactions on Biomedical Engineering (TBME)*, vol. 58, no. 8, pp. 2280–2289, Aug. 2011. 57, 75, 98, 130

[34] D. Cohen, E. Mayer, D. Chen, A. Anstee, J. Vale, G.-Z. Yang, A. Darzi, and P. E. Edwards, "Augmented Reality Image Guidance in Minimally Invasive Prostatectomy," in *Prostate Cancer Imaging Diagnosis, Prognosis, and Intervention*, Beijing, China, 2010, pp. 101–110. 46, 135

[35] A. Cole, S. Ziauddin, and M. Greenspan, "High-speed Imperceptible Structured Light Depth Mapping," in *International Conference on Computer Vision Theory and Applications (VISAPP)*, Valletta, Malta, 2020, pp. 676–684. 75

[36] T. Collins and A. Bartoli, "Towards Live Monocular 3D Laparoscopy Using Shading and Specularity Information," in *Information Processing in Computer-Assisted Interventions (IPCAI)*, Pisa, Italy, 2012, pp. 11–21. 13, 72, 73

[37] T. Collins, D. Pizarro, A. Bartoli, M. Canis, and N. Bourdel, "Realtime Wide-Baseline Registration of the Uterus in Laparoscopic Videos Using Multiple Texture Maps," in *Augmented Reality Environments for Medical Imaging and Computer-Assisted Interventions (AE-CAI)*, Nagoya, Japan, 2013, pp. 162–171. 13, 21, 22, 52, 57, 72, 73, 135, 139

[38] ——, "Computer-Assisted Laparoscopic Myomectomy by Augmenting the Uterus with Pre-Operative MRI Data," in *International Symposium on Mixed and Augmented Reality (ISMAR)*, Munich, Germany, Jul. 2014, pp. 243–248. 21, 22, 23, 53, 139

[39] A. V. Dalca, G. Balakrishnan, J. Guttag, and M. R. Sabuncu, "Unsupervised learning for fast probabilistic diffeomorphic registration," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Granada, Spain, 2018, pp. 729–738. 38

[40] B. D. de Vos, F. F. Berendsen, M. A. Viergever, H. Sokooti, M. Staring, and I. Išgum, "A deep learning framework for unsupervised affine and deformable image registration," *Medical Image Analysis (MIA)*, vol. 52, pp. 128–143, Feb. 2019. 38

[41] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Miami, FL, USA, 2009, pp. 248–255. 50

[42] F. Devernay, F. Mourgues, and E. Coste-Maniere, "Towards Endoscopic Augmented Reality for Robotically Assisted Minimally Invasive Cardiac Surgery," in *International Workshop on Medical Imaging and Augmented Reality (MIAR)*, Shatin, China, 2001, pp. 16–20. 28, 46, 75

[43] P. Dubach, B. Bell, S. Weber, and M. Caversaccio, "Image-Guided Otorhinolaryngology," in *Intraoperative Imaging and Image-Guided Therapy*, F. A. Jolesz, Ed. New York, NY, USA: Springer, 2014, pp. 845–856. 3

[44] N. Dym and S. Z. Kovalsky, "Linearly Converging Quasi Branch and Bound Algorithms for Global Rigid Registration," in *IEEE International Conference on Computer Vision (ICCV)*, Seoul, Korea, 2019, pp. 1628–1636. 139

[45] P. J. Edwards, A. P. King, C. R. J. Maurer, D. A. de Cunha, D. J. Hawkes, D. L. G. Hill, R. P. Gaston, M. R. Fenlon, A. Jusczyzck, A. J. Strong, C. L. Chandler, and M. J. Gleeson, "Design and Evaluation of a System for Microscope-Assisted Guided Interventions (MAGI)," *IEEE Transactions on Medical Imaging (TMI)*, vol. 19, pp. 1082–1093, Nov. 2000. 3

[46] J. Fan, X. Cao, Z. Xue, P.-T. Yap, and D. Shen, "Adversarial similarity network for evaluating image alignment in deep learning based registration," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Granada, Spain, 2018, pp. 739–746. 38

[47] M. Feuerstein, S. M. Wildhirt, R. Bauernschmitt, and N. Navab, "Automatic Patient Registration for Port Placement in Minimally Invasive Endoscopic Surgery," in *Medical Image Computing and Computer Assisted Interventions (MICCAI)*, Palm Springs, USA, 2005, pp. 287–294. 23

[48] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Communications of the ACM (CACM)*, vol. 24, no. 6, pp. 381–395, Jun. 1981. 19, 41, 89

[49] J. M. Fitzpatrick, J. B. West, and C. R. J. Maurer, "Predicting Error in Rigid-Body Point-Based Registration," *IEEE Transactions on Medical Imaging (TMI)*, vol. 17, pp. 694–702, Oct. 1998. 17

[50] A. Fusiello, V. Roberto, and E. Trucco, "Efficient Stereo with Multiple Windowing," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, San Juan, Puerto Rico, 1997, pp. 858–863. 52, 78

[51] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and R. Medina-Carnicer, "Generation of fiducial marker dictionaries using Mixed Integer Linear Programming," *Pattern Recognition*, vol. 51, pp. 481–491, Mar. 2016. 98

[52] A. Geiger, P. Lenz, and R. Urtasun, "Are We Ready For Autonomous Driving? The KITTI Vision Benchmark Suite," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, Providence, Rhode Island, 2012, pp. 3354–3361. 50

[53] N. Gelfand, N. J. Mitra, L. J. Guibas, and H. Pottmann, "Robust Global Registration," *Eurographics Symposium on Geometry Processing (SGP)*, pp. 5–14, 2005. 64, 65

[54] S. Geman and D. E. McClure, "Statistical methods for tomographic image reconstruction," *Bulletin of the International Statistical Institute*, vol. LII-4, pp. 5–21, 1987. 81

[55] J. Geurten, W. Xia, U. Jayarathne, T. M. Peters, and E. C. Chen, "Endoscopic laser surface scanner for minimally invasive abdominal surgeries," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Granada, Spain, 2018, pp. 143–150. 15, 75

[56] Ó. G. Grasa, E. Bernal, S. Casado, I. Gil, and J. M. M. Montiel, "Visual SLAM for Handheld Monocular Endoscope," *IEEE Transactions on Medical Imaging (TMI)*, vol. 33, no. 1, pp. 135–146, Jan. 2014. 13, 21, 22, 23, 72, 73, 131, 135

[57] W. E. L. Grimson, G. J. Ettinger, S. J. White, T. Lozano-Perez, W. M. Wells, III, and R. Kikinis, "An Automatic Method for Frameless Stereotaxy, Image Guided Surgery, and Enhanced Reality Visualization," *IEEE Transactions on Medical Imaging (TMI)*, vol. 15, no. 2, pp. 129–140, Apr. 1996. 3

[58] W. E. L. Grimson, M. Leventon, G. Ettinger, A. Chabrerie, F. Ozlen, S. Nakajima, H. Atsumi, R. Kikinis, and P. Black, "Clinical Experience with a High Precision Image-guided Neurosurgery System," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Cambridge, MA, USA, 1998, pp. 63–73. 3

[59] N. Haouchine, S. Cotin, I. Peterlik, J. Dequidt, M. S. Lopez, E. Kerrien, and M.-O. Berger, "Impact of Soft Tissue Heterogeneity on Augmented Reality for Liver Surgery," *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 21, no. 5, pp. 584–597, May 2015. 48

[60] N. Haouchine, J. Dequidt, I. Peterlik, E. Kerrien, M.-O. Berger, and S. Cotin, "Image-guided simulation of heterogeneous tissue deformation for augmented reality during hepatic surgery," in *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, Adelaide, SA, Australia, 2013, pp. 199–208. 28, 46, 75

[61] C. Harris and M. Stephens, "A Combined Corner and Edge Detection," in *Alvey Vision Conference*, Manchester, England, 1988, pp. 147–151. 82

[62] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge, UK: Cambridge University Press, 2004. 19

[63] G. Haskins, U. Kruger, and P. Yan, "Deep learning in medical image registration: a survey," *arXiv preprint arXiv:1903.02026*, 2019. 38

[64] G. Hattab, M. Arnold, L. Strenger, M. Allan, D. Arsentjeva, O. Gold, T. Simpfendörfer, L. Maier-Hein, and S. Speidel, "Kidney edge detection in laparoscopic image data for computer-assisted surgery," *International Journal of Computer Assisted Radiology and Surgery (IJCARS)*, pp. 1–9, 2019. 38

[65] Y. Hayashi, K. Misawa, D. J. Hawkes, and K. Mori, "Progressive internal landmark registration for surgical navigation in laparoscopic gastrectomy for gastric cancer," *International Journal of Computer Assisted Radiology and Surgery (IJCARS)*, vol. 11, no. 5, pp. 837–845, Jan. 2016. 18, 53, 73

[66] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 2961–2969. 39, 40

[67] M. P. Heinrich, "Closing the gap between deep and conventional image registration using probabilistic dense displacement networks," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Shenzhen, China, 2019, pp. 50–58. 38

[68] A. Hering, B. van Ginneken, and S. Heldmann, "mlvirnet: Multilevel variational image registration network," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Shenzhen, China, 2019, pp. 257–265. 38

[69] A. J. Herline, J. D. Stefansic, J. P. Delbelak, S. L. Hartmann, C. W. Pinson, R. L. Galloway, and W. C. Chapman, "Image-Guided Surgery: Preliminary

Feasibility Studies of Frameless Stereotactic Liver Surgery," *Archives of Surgery*, vol. 134, pp. 644–650, Jun. 1999. 33, 47

[70] D. R. Hess, "Respiratory Mechanics in Mechanically Ventilated Patients," *Respiratory Care*, vol. 59, no. 11, pp. 1773–1794, 2014. 36

[71] H. Hirschmuller, "Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, San Diego, CA, USA, 2005, pp. 807–814. 31, 40

[72] ——, "Stereo Processing by Semiglobal Matching and Mutual Information," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 30, no. 2, pp. 328–341, 2008. 31, 40

[73] B. K. P. Horn, *Robot Vision*. Cambridge, MA, USA: The MIT Press, 1986. 81

[74] M. Hu, G. P. Penney, D. Rueckert, P. J. Edwards, F. Bello, R. Casula, M. Figl, and D. J. Hawkes, "Non-Rigid Reconstruction of the Beating Heart Surface for Minimally Invasive Cardiac Surgery," in *Medical Image Computing and Computer Assisted Interventions (MICCAI)*, London, England, 2009, pp. 34–42. 13, 72, 73

[75] X. Hu, M. Kang, W. Huang, M. R. Scott, R. Wiest, and M. Reyes, "Dual-stream pyramid registration network," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Shenzhen, China, 2019, pp. 382–390. 38

[76] Y. Hu, M. Modat, E. Gibson, W. Li, N. Ghavami, E. Bonmati, G. Wang, S. Bandula, C. M. Moore, M. Emberton, S. Ourselin, J. A. Noble, D. C. Barratt, and T. Vercauteren, "Weakly-supervised convolutional neural networks for multimodal image registration," *Medical Image Analysis (MIA)*, vol. 49, pp. 1–13, 2018. 38

[77] R. Hussain, A. Lalande, R. Marroquin, K. B. Girum, C. Guigou, and A. B. Grayeli, "Real-time augmented reality for ear surgery," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Granada, Spain, 2018, pp. 324–331. 17, 18, 73

[78] S. Ieiri, M. Uemura, K. Konishi, R. Souzaki, Y. Nagao, N. Tsutsumi, T. Akahoshi, K. Ohuchida, T. Ohdaira, M. Tomikawa, K. Tanoue, M. Hashizume, and T. Taguchi, "Augmented reality navigation system for laparoscopic splenectomy in children based on preoperative CT image using optical tracking device," *Pediatric Surgery International (PSI)*, vol. 28, no. 4, pp. 341–346, 2012. 17

[79] Intuitive Surgical, "Intuitive Surgical, Inc. - da Vinci Surgical System," https://www.intuitivesurgical.com/, 2017, accessed: 2017-06-29. 32

[80] F. Isensee, J. Petersen, S. A. Kohl, P. F. Jäger, and K. H. Maier-Hein, "nnU-Net: Breaking the Spell on Successful Medical Image Segmentation," *arXiv preprint arXiv:1904.08128*, 2019. 38

[81] H. Jin, P. Favaro, and S. Soatto, "Real-time feature tracking and outlier rejection with changes in illumination," in *IEEE International Conference on Computer Vision (ICCV)*, Vancouver, BC, Canada, 2001, pp. 684–689. 20

[82] A. E. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 21, no. 5, pp. 433–449, May 1999. 64, 65

[83] I. T. Jolliffe, *Principle Component Analysis*, 2nd ed., ser. Springer Series in Statistics. New York, USA: Springer-Verlag, 2002. 42

[84] W. Kay, J. . Carreira, K. Simonyan, B. Zhang, C. Hillier, S. Vijayanarasimhan, F. Viola, T. Back, P. Natsev, M. Suleyman, and A. Zisserman, "The Kinetics Human Action Video Dataset," *arXiv preprint arXiv:1705.06950*, pp. 1–22, May 2017. 50

[85] T. P. Kingham, S. Jayaraman, L. W. Clements, M. A. Scherer, J. D. Stefansic, and W. R. Jarnagin, "Evolution of Image-Guided Liver Surgery: Transition from Open to Laparoscopic Procedures," *Journal of Gastrointestinal Surgery (JOGS)*, vol. 17, no. 7, pp. 1274–1282, May 2013. 51

[86] T. P. Kingham, L. M. Pak, A. L. Simpson, U. Leung, A. Doussot, M. I. D'Angelica, R. P. DeMatteo, P. J. Allen, and W. R. Jarnagin, "3D image guidance assisted identification of colorectal cancer liver metastases not seen on intraoperative ultrasound: results from a prospective trial," *HPB : The Official Journal of the International Hepato Pancreato Biliary Association*, vol. 20, no. 3, pp. 260–267, 2017. 4

[87] T. P. Kingham, M. A. Scherer, B. W. Neese, L. W. Clements, J. D. Stefansic, and W. R. Jarnagin, "Image-guided liver surgery: intraoperative projection of computed tomography images utilizing tracked ultrasound," *HPB : The*

Official Journal of the International Hepato Pancreato Biliary Association, vol. 14, no. 9, pp. 594–603, Jan. 2012. 4, 51, 132

[88] L. Kneip, D. Scaramuzza, and R. Siegwart, "A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, CO, USA, 2011, pp. 2969–2976. 25

[89] J. Krebs, T. Mansi, H. Delingette, L. Zhang, F. C. Ghesu, S. Miao, A. K. Maier, N. Ayache, R. Liao, and A. Kamen, "Robust non-rigid registration through agent-based action learning," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Québec City, QC, Canada, 2017, pp. 344–352. 38

[90] J. Lamarca, S. Parashar, A. Bartoli, and J. M. M. Montiel, "DefSLAM: Tracking and Mapping of Deforming Scenes from Monocular Sequences," *arXiv preprint arXiv:1908.08918*, 2019. 13, 21, 22, 72, 73

[91] P. Lamata, W. Ali, A. Cano, J. Cornella, J. Declerck, O. J. Elle, A. Freudenthal, H. Furtado, D. Kalkofen, E. Naerum, E. Samset, P. Sánchez-Gonzalez, F. M. Sánchez-Margallo, D. Schmalstieg, M. Sette, T. Stüdeli, J. Vander Sloten, and E. J. Gómez, "Augmented Reality for Minimally Invasive Surgery: Overview and Some Recent Advances," in *Augmented Reality*, S. Maad, Ed. Croatia: INTECH Open Access Publisher, 2010, pp. 73–98. 9, 46, 57

[92] A. H. Land and A. G. Doig, "An Automatic Method of Solving Discrete Programming Problems," *Econometrica*, vol. 28, pp. 497–520, Jul. 1960. 66

[93] T. Lange, M. Hünerbein, S. Eulenstein, S. Beller, and P. M. Schlag, "Development of Navigation Systems for Image-Guided Laparoscopic Tumor Resections in Liver Surgery," *Recent Results in Cancer Research*, vol. 167, pp. 13–36, Sep. 2005. 117, 131

[94] W. W. Lau, N. A. Ramey, J. J. Corso, N. V. Thakor, and G. D. Hager, "Stereo-Based Endoscopic Tracking of Cardiac Surface Deformation," in *Medical Image Computing and Computer Assisted Interventions (MICCAI)*, St. Malo, France, 2004, pp. 494–501. 27, 76

[95] M. C. Lee, O. Oktay, A. Schuh, M. Schaap, and B. Glocker, "Image-and-spatial transformer networks for structure-guided image registration," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Shenzhen, China, 2019, pp. 337–345. 38

[96] S.-L. Lee, M. Lerotic, V. Vitiello, S. Giannarou, K.-W. Kwok, M. Visentini-Scarzanella, and G.-Z. Yang, "From Medical Images to Minimally Invasive Intervention: Computer Assistance for Robotic Surgery," *Computerized Medical Imaging and Graphics (CMIG)*, vol. 34, no. 1, pp. 33–45, 2010. 9, 57

[97] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EPnP: An Accurate O(n) Solution to the PnP Problem," *International Journal of Computer Vision (IJCV)*, vol. 81, no. 2, pp. 155–166, Jul. 2008. 19

[98] U. Leung, A. L. Simpson, L. B. Adams, W. R. Jarnagin, M. I. Miga, and T. P. Kingham, "Image guidance improves localization of sonographically occult colorectal liver metastases," in *SPIE Medical Imaging 2015: Image-Guided*

Procedures, Robotic Interventions, and Modeling, Orlando, FL, USA, Mar. 2015, p. 94151L. 4

[99] R. Liao, S. Miao, P. de Tournemire, S. Grbic, A. Kamen, T. Mansi, and D. Comaniciu, "An artificial agent for robust image registration," in *AAAI Conference on Artificial Intelligence (AAAI)*, San Francisco, CA, USA, 2017. 38

[100] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common Objects in context," in *European Conference on Computer Vision (ECCV)*, Zürich, Switzerland, 2014, pp. 740–755. 41

[101] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis (MIA)*, vol. 42, pp. 60–88, 2017. 38

[102] W. P. Liu, J. D. Richmon, J. M. Sorger, M. Azizian, and R. H. Taylor, "Augmented reality and cone beam CT guidance for transoral robotic surgery," *Journal of Robotic Surgery (JRS)*, vol. 9, no. 3, pp. 223–233, Jul. 2015. 3

[103] X. Liu, A. Sinha, M. Ishii, G. D. Hager, A. Reiter, R. H. Taylor, and M. Unberath, "Self-supervised learning for dense depth estimation in monocular endoscopy," *arXiv preprint arXiv:1902.07766*, 2019. 38, 72, 73

[104] X. Liu, A. Sinha, M. Ishii, G. D. Hager, R. H. Taylor, and M. Unberath, "Self-supervised Dense 3D Reconstruction from Monocular Endoscopic Video," *arXiv preprint arXiv:1909.03101*, 2019. 39, 54

[105] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015, pp. 3431–3440. 39, 41

[106] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision (IJCV)*, vol. 60, no. 2, pp. 91–110, 2004. 20, 83

[107] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift für Medizinische Physik*, vol. 29, no. 2, pp. 102–127, 2019. 38

[108] X. Luo, Y. Wan, H.-Q. Zeng, Y. Guo, H. C. Ewurum, X.-B. Zhang, A. J. McLeod, and T. M. Peters, "Simultaneous surgical visibility assessment, restoration, and augmented stereo surface reconstruction for robotic prostatectomy," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Granada, Spain, 2018, pp. 315–323. 27

[109] B. Ma, T. P. Kingham, M. I. Miga, W. R. Jarnagin, and A. L. Simpson, "Liver segmentation in color images," in *SPIE Medical Imaging 2017: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 10135, Orlando, FL, USA, 2017, p. 101351O. 38

[110] R. Ma, R. Wang, S. Pizer, J. Rosenman, S. K. McGill, and J.-M. Frahm, "Real-time 3d reconstruction of colonoscopic surfaces for determining missing regions," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Shenzhen, China, 2019, pp. 573–582. 23, 39, 54

[111] D. J. C. MacKay, *Information theory, inference and learning algorithms*. New York, NY, USA: Cambridge University Press, 2003. 19

[112] N. Mahmoud, Ó. G. Grasa, S. A. Nicolau, C. Doignon, L. Soler, J. Marescaux, and J. M. M. Montiel, "On-patient see-through augmented reality based on visual SLAM," *International Journal of Computer Assisted Radiology and Surgery (IJCARS)*, vol. 12, no. 1, pp. 1–11, 2017. 21, 22

[113] L. Maier-Hein, P. Mountney, A. Bartoli, H. Elhawary, D. Elson, A. Groch, A. Kolb, M. Rodrigues, J. Sorger, S. Speidel, and D. Stoyanov, "Optical techniques for 3D surface reconstruction in computer-assisted laparoscopic surgery," *Medical Image Analysis (MIA)*, vol. 17, no. 8, pp. 974–996, Dec. 2013. 9, 28, 29, 47, 48, 53, 57, 135

[114] A. Makadia, A. Patterson, IV, and K. Daniilidis, "Fully Automatic Registration of 3D Point Clouds," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, New York, NY, USA, 2006, pp. 1297–1304. 65

[115] A. Malti, A. Bartoli, and T. Collins, "Template-Based Conformal Shape-from-Motion from Registered Laparoscopic Images." *Medical Image Understanding and Analysis (MIUA)*, pp. 1–5, 2011. 13, 22, 72, 73

[116] ——, "Template-Based Conformal Shape-from-Motion-and-Shading for Laparoscopy," in *International Conference on Information Processing in Computer-Assisted Interventions (IPCAI)*, Pisa, Italy, 2012, pp. 1–10. 13, 72, 73

[117] S. Mattoccia, S. Giardino, and A. Gambini, "Accurate and Efficient Cost Aggregation Strategy for Stereo Correspondence Based on Approximated

Joint Bilateral Filtering," in *Asian Conference of Computer Vision (ACCV)*, Xi'an, China, 2009, pp. 371–380. 31

[118] C. R. J. Maurer, J. M. Fitzpatrick, M. Y. Wang, R. L. J. Galloway, R. J. Maciunas, and G. S. Allen, "Registration of Head Volume Images Using Implantable Fiducial Markers," *IEEE Transactions on Medical Imaging (TMI)*, vol. 16, pp. 447–462, Aug. 1997. 17

[119] C. R. J. Maurer, R. J. Maciunas, and J. M. Fitzpatrick, "Registration of Head CT Images to Physical Space Using a Weighted Combination of Points and Surfaces," *IEEE Transactions on Medical Imaging (TMI)*, vol. 17, pp. 753–761, Oct. 1998. 17

[120] K. Mikolajczyk and C. Schmid, "An Affine Invariant Interest Point Detector," in *European Conference on Computer Vision (ECCV)*, Copenhagen, Denmark, 2002, pp. 128–142. 49

[121] D. J. Mirota, M. Ishii, and G. D. Hager, "Vision-Based Navigation in Image-Guided Interventions," *Annual Review of Biomedical Engineering*, vol. 13, no. 1, pp. 297–319, Aug. 2011. 9, 57

[122] D. J. Mirota, A. Uneri, S. Schafer, S. Nithiananthan, D. D. Reh, M. Ishii, G. L. Gallia, R. H. Taylor, G. D. Hager, and J. H. Siewerdsen, "Evaluation of a System for High-Accuracy 3D Image-Based Registration of Endoscopic Video to C-Arm Cone-Beam CT for Image-Guided Skull Base Surgery," *IEEE Transactions on Medical Imaging (TMI)*, vol. 32, no. 7, pp. 1215–1226, Jul. 2013. 3, 33

[123] D. J. Mirota, H. Wang, R. H. Taylor, M. Ishii, G. L. Gallia, and G. D. Hager, "A System for Video-Based Navigation for Endoscopic Endonasal Skull Base Surgery," *IEEE Transactions on Medical Imaging (TMI)*, vol. 31, no. 4, pp. 963–976, 2012. 13, 18, 72, 73, 131

[124] M. Moghbel, S. Mashohor, R. Mahmud, and M. I. B. Saripan, "Review of liver segmentation and computer assisted detection/diagnosis methods in computed tomography," *Artificial Intelligence Review*, vol. 50, no. 4, pp. 497–537, 2018. 38

[125] P. Mountney, J. Fallert, S. Nicolau, L. Soler, and P. W. Mewes, "An Augmented Reality Framework for Soft Tissue Surgery," in *Medical Image Computing and Computer Assisted Interventions (MICCAI)*, Cambridge, MA, USA, 2014, pp. 423–431. 16, 33, 34, 46, 48

[126] P. Mountney and G.-Z. Yang, "Soft Tissue Tracking for Minimally Invasive Surgery: Learning Local Deformation Online," in *Medical Image Computing and Computer Assisted Interventions (MICCAI)*, New York, NY, USA, 2008, pp. 364–372. 36, 37

[127] ——, "Motion Compensated SLAM for Image Guided Surgery," in *Medical Image Computing and Computer Assisted Interventions (MICCAI)*, Beijing, China, 2010, pp. 496–504. 37, 131

[128] S. Nicolau, L. Soler, D. Mutter, and J. Marescaux, "Augmented reality in laparoscopic surgical oncology," *Surgical Oncology (SO)*, vol. 20, no. 3, pp. 189–201, 2011. 9, 57

[129] S. A. Nicolau, M. Diana, V. Agnus, L. Soler, and J. Marescaux, "Semi-Automated Augmented Reality for Laparoscopic Surgery: First In-vivo Evaluation," *International Journal of Computer Assisted Radiology and Surgery (IJCARS)*, vol. 8, no. S1, pp. 109–110, May 2013. 18, 53, 73, 135

[130] D. P. Noonan, P. Mountney, D. S. Elson, A. Darzi, and G.-Z. Yang, "A stereoscopic fibroscope for camera motion and 3D depth recovery during Minimally Invasive Surgery," in *IEEE International Conference on Robotics and Automation (ICRA)*, Kobe, Japan, 2009, pp. 4463–4468. 13, 72, 73, 131, 135

[131] Northern Digital Inc., "Polaris Series Optical Tracking System," https://www.ndigital.com/msci/products/polaris-series, 2017, accessed: 2017-07-31. 132, 193

[132] O. Oktay, L. Zhang, T. Mansi, P. Mountney, P. Mewes, S. A. Nicolau, L. Soler, and C. Chefd'hotel, "Biomechanically Driven Registration of Pre- to Intra-Operative 3D Images for Laparoscopic Surgery," in *Medical Image Computing and Computer Assisted Interventions (MICCAI)*, Nagoya, Japan, 2013, pp. 1–9. 35

[133] L. M. Pak, J. Gagnière, P. J. Allen, V. P. Balachandran, M. I. D'Angelica, R. P. DeMatteo, W. R. Jarnagin, M. I. Miga, A. L. Simpson, and T. P. Kingham, "Utility of Image Guidance in the Localization of Disappearing Colorectal Liver Metastases," *Journal of Gastrointestinal Surgery (JOGS)*, vol. 23, pp. 760–767, 2019. 4

[134] F. Panaro, H. Habibeh, P. Pessaux, and F. Navarro, "Navigation Liver Surgery for Complex Hydatid Cyst with Biliary Tree Communication," *International Journal of Surgery (IJS) Case Reports*, vol. 12, pp. 112–116, 2015. 4

[135] C. Papazov and D. Burschka, "Stochastic Global Optimization for Robust Point Set Registration," *Computer Vision and Image Understanding (CVIU)*, vol. 115, no. 12, pp. 1598–1609, Dec. 2011. 64

[136] S. Paris, P. Kornprobst, J. Tumblin, and F. Durand, "A Gentle Introduction to Bilateral Filtering and Its Applications," in *Special Interest Group on Computer Graphics and Interactive Techniques (SIGGRAPH) Courses*, San Diego, CA, USA, 2007, pp. 1–45. 31

[137] M. Peterhans, A. vom Berg, B. Dagon, D. Inderbitzin, C. Baur, D. Candinas, and S. Weber, "A navigation system for open liver surgery: design, workflow and first clinical applications," *International Journal of Medical Robotics and Computer Assisted Surgery (IJMRCAS)*, vol. 7, no. 1, pp. 7–16, Oct. 2010. 4, 132

[138] T. M. Peters, "Image-guidance for surgical procedures," *Physics in Medicine and Biology (PMB)*, vol. 51, no. 14, pp. R505–R540, Jun. 2006. 1, 17

[139] R. Plantefève, N. Haouchine, J. P. Radoux, and S. Cotin, "Automatic Alignment of pre and Intraoperative Data using Anatomical Landmarks for Augmented Laparoscopic Liver Surgery," in *International Symposium on Biomedical Simulation (ISBMS)*, Strasbourg, France, 2014, pp. 58–66. 53

[140] R. Plantefève, I. Peterlik, N. Haouchine, and S. Cotin, "Patient-Specific Biomechanical Modeling for Guidance During Minimally-Invasive Hepatic

Surgery," *Annals of Biomedical Engineering (ABME)*, vol. 44, no. 1, pp. 139–153, Aug. 2015. 116, 131

[141] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge, MA, USA: Cambridge University Press, 1988. 81

[142] G. A. Puerto-Souza, J. A. Cadeddu, and G. L. Mariottini, "Toward Long-Term and Accurate Augmented-Reality for Monocular Endoscopic Videos," *IEEE Transactions on Biomedical Engineering (TBME)*, vol. 61, no. 10, pp. 2609–2620, 2014. 17, 18, 46, 73

[143] G. A. Puerto-Souza and G. L. Mariottini, "A Fast and Accurate Feature-Matching Algorithm for Minimally-Invasive Endoscopic Images," *IEEE Transactions on Medical Imaging (TMI)*, vol. 32, no. 7, pp. 1201–1214, Jul. 2013. 20, 53

[144] H. Ravishankar, R. Venkataramani, S. Anamandra, P. Sudhakar, and P. Annangi, "Feature transformers: Privacy preserving lifelong learners for medical imaging," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Shenzhen, China, 2019, pp. 347–355. 55

[145] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *Advances in Neural Information Processing Systems (NIPS)*, Montréal, QC, Canada, 2015, pp. 91–99. 41

[146] R. Richa, A. P. Bó, and P. Poignet, "Towards Robust 3D Visual Tracking for Motion Compensation in Beating Heart Surgery," *Medical Image Analysis (MIA)*, vol. 15, no. 3, pp. 302–315, Jun. 2011. 49

[147] D. W. Roberts, J. W. Strohbehn, J. F. Hatch, W. Murray, and H. Kettenberger, "A frameless stereotaxic integration of computerized tomographic imaging and the operating microscope," *Journal of Neurosurgery (JNS)*, vol. 65, pp. 545–549, Oct. 1986. 3

[148] C. Robertson and R. B. Fisher, "Parallel Evolutionary Registration of Range Data," *Computer Vision and Image Understanding (CVIU)*, vol. 87, no. 1, pp. 39–50, Jul. 2002. 64

[149] M. R. Robu, J. Ramalhinho, S. Thompson, K. Gurusamy, B. Davidson, D. Hawkes, D. Stoyanov, and M. J. Clarkson, "Global rigid registration of CT to video in laparoscopic liver surgery," *International Journal of Computer Assisted Radiology and Surgery (IJCARS)*, vol. 13, no. 6, pp. 947–956, May 2018. 53, 129, 139

[150] M.-M. Rohé, M. Datar, T. Heimann, M. Sermesant, and X. Pennec, "Svf-net: Learning deformable image registration using shape matching," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Québec City, QC, Canada, 2017, pp. 266–274. 38

[151] S. Röhl, S. Bodenstedt, S. Suwelack, H. Kenngott, B. P. Mueller-Stich, R. Dillmann, and S. Speidel, "Real-Time Surface Reconstruction from Stereo Endoscopic Images for Intraoperative Registration," in *SPIE Medical Imaging 2011: Visualization, Image-Guided Procedures, and Modeling*, Lake Buena

Vista, FL, USA, 2011, pp. 796 414–1–796 414–13. 23, 28, 29, 30, 32, 46, 49, 75, 76

[152] S. Röhl, S. Speidel, D. Gonzalez-Aguirre, S. Suwelack, H. Kenngott, T. Asfour, B. P. Muller-Stich, and R. Dillmann, "From Stereo Image Sequences to Smooth and Robust Surface Models using Temporal Information and Bilateral Postprocessing," in *International Conference on Robotics and Biomimetics (ROBIO)*, Phucket, Thailand, 2011, pp. 544–550. 28, 29, 31, 46, 49, 75, 76

[153] S. Röhl, S. Bodenstedt, S. Suwelack, H. Kenngott, B. P. Müller-Stich, R. Dillmann, and S. Speidel, "Dense GPU-Enhanced Surface Reconstruction from Stereo Endoscopic Images for Intraoperative Registration," *Medical Physics (MP)*, vol. 39, no. 3, pp. 1632–1645, Mar. 2012. 28, 29, 31, 32, 46, 49, 75, 76

[154] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Munich, Germany, 2015, pp. 234–241. 38

[155] C. B. Rosen, "Management of Hepatic Metastases," *Cancer Control (CCX)*, vol. 5, pp. 30–31, 1998. 3

[156] E. Rosten and T. Drummond, "Machine Learning for High-Speed Corner Detection," in *European Conference on Computer Vision (ECCV)*, Graz, Austria, 2006, pp. 430–443. 82

[157] E. Rublee, V. Rabaud, K. Konolige, and G. R. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *IEEE International Conference on Computer Vision (ICCV)*, Barcelona, Spain, 2011, p. 25642571. 25, 82

[158] D. C. Rucker, Y. Wu, L. W. Clements, J. E. Ondrake, T. S. Pheiffer, A. L. Simpson, W. R. Jarnagin, and M. I. Miga, "A Mechanics-Based Nonrigid Registration Method for Liver Surgery Using Sparse Intraoperative Data," *IEEE Transactions on Medical Imaging (TMI)*, vol. 33, no. 1, pp. 147–158, Jan. 2014. 15, 47, 54, 117, 132, 141

[159] R. B. Rusu, N. Blodow, and M. Beetz, "Fast Point Feature Histograms (FPFH) for 3D Registration," in *IEEE International Conference on Robotics and Automation (ICRA)*. Kobe, Japan: IEEE, 2009, pp. 3212–3217. 64, 65

[160] R. B. Rusu, Z. C. Marton, N. Blodow, M. Dolha, and M. Beetz, "Towards 3D Point cloud based object maps for household environments," *Robotics and Autonomous Systems*, vol. 56, no. 11, pp. 927–941, Nov. 2008. 80

[161] R. Sandhu, S. Dambreville, and A. Tannenbaum, "Point Set Registration via Particle Filtering and Stochastic Dynamics," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 32, no. 8, pp. 1459–1473, Aug. 2010. 64

[162] H. Sarbolandi, D. Lefloch, and A. Kolb, "Kinect range sensing: Structured-light versus Time-of-Flight Kinect ," *Computer Vision and Image Understanding (CVIU)*, vol. 139, pp. 1–20, Oct. 2015. 74

[163] A. Sardi and A. Akbarov, "Management of Primary and Metastatic Tumors to the Liver," *Oncology*, vol. 10, pp. 911–938, 1996. 3

[164] F. Schaffalitzky and A. Zisserman, "Multi-view Matching for Unordered Image Sets, or "How Do I Organize My Holiday Snaps?"," in *European*

Conference on Computer Vision (ECCV), Copenhagen, Denmark, 2002, pp. 414–431. 49

[165] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," *International Journal of Computer Vision (IJCV)*, vol. 47, pp. 7–42, 2002. 50, 116

[166] L. V. Selby, R. P. DeMatteo, R. M. Tholey, W. R. Jarnagin, J. Garcia-Aguilar, P. D. Strombom, P. J. Allen, T. P. Kingham, M. R. Weiser, M. F. Brennan, and V. E. Strong, "Evolving application of minimally invasive cancer operations at a tertiary cancer center," *Journal of Surgical Oncology (JSO)*, vol. 115, no. 4, pp. 365–370, Nov. 2016. 9, 57, 133

[167] T. Sentker, F. Madesta, and R. Werner, "GDL-FIRE^{4D}: Deep Learning-Based Fast 4D CT Image Registration," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Granada, Spain, 2018, pp. 765–773. 38

[168] J. Shi and C. Tomasi, "Good Features to Track," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 1994, pp. 593–600. 19, 82, 84, 85, 93

[169] J. H. Shuhaiber, "Augmented Reality in Surgery," *Archives of Surgery*, vol. 139, no. 2, pp. 170–174, 2004. 46

[170] L. Silva, O. R. P. Bellon, and K. L. Boyer, "Precision Range Image Registration Using a Robust Surface Interpenetration Measure and Enhanced Genetic Algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 27, no. 5, pp. 762–776, May 2005. 64

[171] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014. 39

[172] T. Simpfendörfer, M. Baumhauer, M. Müller, C. N. Gutt, H.-P. Meinzer, J. J. Rassweiler, S. Guven, and T. Doguu, "Augmented Reality Visualization During Laparoscopic Radical Prostatectomy," *Journal of Endourology*, vol. 25, no. 12, pp. 1841–1845, Dec. 2011. 17, 18, 46, 47, 73

[173] A. L. Simpson, P. Dumpuri, W. R. Jarnagin, and M. I. Miga, "Model-Assisted Image-Guided Liver Surgery Using Sparse Intraoperative Data," *Soft Tissue Biomechanical Modeling for Computer Assisted Surgery*, vol. 11, pp. 7–40, 2012. 15, 33, 47, 54, 57, 87, 93, 136, 141

[174] A. L. Simpson, D. A. Geller, A. W. Hemming, W. R. Jarnagin, L. W. Clements, M. I. D'Angelica, P. Dumpuri, M. Gönen, I. Zendejas, M. I. Miga, and J. D. Stefansic, "Liver Planning Software Accurately Predicts Postoperative Liver Volume and Measures Early Regeneration," *Journal of the American College of Surgeons (JACS)*, vol. 219, no. 2, pp. 199–207, Aug. 2014. 4

[175] A. L. Simpson and T. P. Kingham, "Current Evidence in Image-Guided Liver Surgery," *Journal of Gastrointestinal Surgery (JOGS)*, vol. 20, no. 6, pp. 1265–1269, 2016. 4, 15, 52, 75, 132

[176] A. Sinha, X. Liu, A. Reiter, M. Ishii, G. D. Hager, and R. H. Taylor, "Endoscopic navigation in the absence of ct imaging," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Granada, Spain, 2018, pp. 64–71. 23

[177] M. Sizintsev, "A Review of Stereo and Motion Integration: Stereomotion," York University, Toronto, Tech. Rep. CSE-2009-06, Oct. 2009. 49

[178] M. Sizintsev, S. Kuthirummal, S. Samarasekera, R. Kumar, H. S. Sawhney, and A. Chaudhry, "GPU Accelerated Realtime Stereo for Augmented Reality," in *International Symposium 3D Data Processing, Visualization and Transmission (3DPVT)*, Paris, France, 2010, pp. 1–8. 92

[179] M. Sizintsev and R. P. Wildes, "Coarse-to-fine stereo vision with accurate 3D boundaries," *Image and Vision Computing (IVC)*, vol. 28, no. 3, pp. 352–366, Mar. 2010. 7, 52, 58, 77, 93, 134

[180] H. Sokooti, B. De Vos, F. Berendsen, B. P. Lelieveldt, I. Išgum, and M. Staring, "Nonrigid image registration using multi-scale 3d convolutional neural networks," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Québec City, QC, Canada, 2017, pp. 232–239. 38

[181] A. D. Speers, B. Ma, W. R. Jarnagin, S. Himidan, A. L. Simpson, and R. P. Wildes, "Fast and accurate vision-based stereo reconstruction and motion estimation for image-guided liver surgery," *Healthcare Technology Letters: Medical Image Computing and Computer Assisted Interventions Workshop (MICCAIW) on Augmented Environments for Computer-Assisted Interventions (AE-CAI)*, vol. 5, no. 5, pp. 208–214, 2018. 9

[182] D. Stoyanov, A. Darzi, and G.-Z. Yang, "A Practical Approach Towards Accurate Dense 3D Depth Recovery for Robotic Laparoscopic Surgery," *Computer Aided Surgery (JCAS)*, vol. 10, no. 4, pp. 199–208, 2010. 49, 76

[183] D. Stoyanov, M. V. Scarzanella, P. Pratt, and G.-Z. Yang, "Real-Time Stereo Reconstruction in Robotically Assisted Minimally Invasive Surgery," in *Medical Image Computing and Computer Assisted Interventions (MICCAI)*, Beijing, China, 2010, pp. 275–282. 27, 36, 76

[184] G. Strang, *Linear Algebra and Its Applications*. New York, NY, USA: Academic Press, 1976. 81

[185] L. M. Su, B. P. Vagvolgyi, R. Agarwal, C. E. Reiley, and R. H. Taylor, "Augmented Reality During Robot-assisted Laparoscopic Partial Nephrectomy: Toward Real-Time 3D-CT to Stereoscopic Video Registration," *Urology*, vol. 73, pp. 896–900, 2009. 28, 45, 75, 135

[186] D. Sun, J. Liu, C. A. Linte, H. Duan, and R. A. Robb, "Surface Reconstruction from Tracked Endoscopic Video Using the Structure from Motion Approach," in *Augmented Reality Environments for Medical Imaging and Computer-Assisted Interventions (AE-CAI)*, Nagoya, Japan, 2013, pp. 127–135. 22

[187] D. Teber, S. Guven, T. Simpfendörfer, M. Baumhauer, E. O. Güven, F. Yencilek, A. S. Gözen, and J. Rassweiler, "Augmented Reality: A New Tool To Improve Surgical Accuracy during Laparoscopic Partial Nephrectomy? Preliminary In Vitro and In Vivo Results," *European Urology*, vol. 56, no. 2, pp. 332–338, Aug. 2009. 17, 18, 46, 47, 73

[188] S. Thompson, J. Totz, Y. Song, S. Johnsen, D. Stoyanov, S. Ourselin, K. Gurusamy, C. Schneider, B. Davidson, D. Hawkes, and M. J. Clarkson, "Accuracy validation of an image guided laparoscopy system for liver resection," in *SPIE*

Medical Imaging 2015: Image-Guided Procedures, Robotic Interventions, and Modeling, vol. 9415, Orlando, FL, USA, 2015, p. 941509. 15, 28, 29, 75

[189] J. Totz, S. Thompson, D. Stoyanov, K. Gurusamy, B. R. Davidson, D. J. Hawkes, and M. J. Clarkson, "Fast Semi-dense Surface Reconstruction from Stereoscopic Video in Laparoscopic Surgery," in *International Conference on Information Processing in Computer-Assisted Interventions (IPCAI)*, Fukoka, Japan, 2014, pp. 206–215. 52, 134

[190] V. C. K. Tse, M. Y. S. Kalani, and J. R. Adler, "Techniques of stereotactic localization," in *Principles and Practice of Stereotactic Radiosurgery*, L. S. Chin and W. F. Regine, Eds. New York, NY, USA: Springer, 2015, pp. 25–32. 2

[191] C. Tsui, R. Klein, and M. Garabrant, "Minimally invasive surgery: national trends in adoption and future directions for hospital strategy," *Surgical Endoscopy*, vol. 27, no. 7, pp. 2253–2257, May 2013. 9, 57

[192] N. Tsutsumi, M. Tomikawa, M. Uemura, T. Akahoshi, Y. Nagao, K. Konishi, S. Ieiri, J. Hong, Y. Maehara, and M. Hashizume, "Image-guided laparoscopic surgery in an open MRI operating theater," *Surgical Endoscopy*, vol. 27, no. 6, pp. 2178–2184, 2013. 17

[193] B. Vagvolgyi, L. M. Su, R. Taylor, and G. D. Hager, "Video to CT Registration for Image Overlay on Solid Organs," in *Augmented Reality in Medical Imaging and Augmented Reality in Computer-Aided Surgery (AMIARCS)*, New York, NY, USA, 2008, pp. 78–86. 28, 29, 43, 44, 45, 49, 75, 76, 135

[194] R. S. Varga, *Matrix Iterative Analysis*. Englewood Cliffs, NJ, USA: PrenticeHall, 1962. 81

[195] J. Vollmer, R. Mencl, and H. Mueller, "Improved Laplacian Smoothing of Noisy Surface Meshes," *Computer Graphics Forum (CGF)*, vol. 18, no. 3, pp. 131–138, 1999. 25

[196] M. P. Wachowiak, R. Smolíková, Y. Zheng, J. M. Zurada, and A. S. Elmaghraby, "An Approach to Multimodal Biomedical Image Registration Utilizing Particle Swarm Optimization," *IEEE Transactions on Evolutionary Computation (TEVC)*, vol. 8, no. 3, pp. 289–301, Jun. 2004. 64

[197] H. Wang, D. Mirota, M. Ishii, and G. D. Hager, "Robust Motion Estimation and Structure Recovery from Endoscopic Image Sequences With an Adaptive Scale Kernel Consensus Estimator," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, Anchorage, AK, USA, 2008, pp. 1–7. 13, 72, 73

[198] H. Wang and C. Schmid, "Action Recognition with Improved Trajectories," in *IEEE International Conference on Computer Vision (ICCV)*. Sydney, Australia: IEEE, 2013, pp. 3551–3558. 83, 84

[199] J. Wang, H. Suenaga, K. Hoshi, L. Yang, E. Kobayashi, I. Sakuma, and H. Liao, "Augmented reality navigation with automatic marker-free image registration using 3-d image overlay for dental surgery," *IEEE Transactions on Biomedical Engineering (BME)*, vol. 61, no. 4, pp. 1295–1304, 2014. 42

[200] R. Wang, S. M. Pizer, and J.-M. Frahm, "Recurrent neural network for (un-) supervised learning of monocular video visual odometry and depth,"

in *Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 5555–5564. 39

[201] C. Wengert, P. Cattin, J. M. Duff, C. Baur, and G. Szekely, "Markerless Endoscopic Registration and Referencing," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Copenhagen, 2006, pp. 816–823. 3

[202] A. R. Widya, Y. Monno, M. Okutomi, S. Suzuki, T. Gotoda, and K. Miki, "Whole Stomach 3D Reconstruction and Frame Localization From Monocular Endoscope Video," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 7, pp. 1–10, 2019. 22, 23, 39, 40, 54

[203] E. Wild, D. Teber, D. Schmid, T. Simpfendörfer, M. Müller, A.-C. Baranski, K. Kopka, and L. Maier-Hein, "Robust augmented reality guidance with fluorescent markers in laparoscopic surgery," *International Journal of Computer Assisted Radiology and Surgery (IJCARS)*, vol. 11, no. 6, pp. 899–907, 2016. 17, 47

[204] C. Winne, M. Khan, F. Stopp, E. Jank, and E. Keeve, "Overlay visualization in endoscopic ENT surgery," *International Journal of Computer Assisted Radiology and Surgery (IJCARS)*, vol. 6, no. 3, pp. 401–406, 2011. 3

[205] O. J. Woodford, M.-T. Pham, A. Maki, F. Perbet, and B. Stenger, "Demisting the Hough Transform for 3D Shape Recognition and Registration," *International Journal of Computer Vision (IJCV)*, vol. 106, no. 3, pp. 332–341, Feb. 2014. 65

[206] C.-H. Wu, Y.-C. Chen, C.-Y. Liu, C.-C. Chang, and Y.-N. Sun, "Automatic Extraction and Visualization of Human Inner Structures from Endoscopic Image Sequences," in *SPIE Medical Imaging 2004: Physiology, Function, and Structure from Medical Images*, San Diego, CA, USA, 2004, pp. 464–473. 22

[207] C.-H. Wu, Y.-N. Sun, and C.-C. Chang, "Three-Dimensional Modeling From Endoscopic Video Using Geometric Constraints Via Feature Positioning," *IEEE Transactions on Biomedical Engineering (TBME)*, vol. 54, no. 7, pp. 1199–1211, 2007. 13, 21, 22, 23, 72, 73

[208] J. Yang, H. Li, D. Campbell, and Y. Jia, "Go-ICP: A Globally Optimal Solution to 3D ICP Point-Set Registration," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 38, no. 11, pp. 2241–2254, 2016. 65, 67, 92, 128, 133

[209] X. Yang, R. Kwitt, M. Styner, and M. Niethammer, "Quicksilver: Fast predictive image registration–a deep learning approach," *NeuroImage*, vol. 158, pp. 378–396, 2017. 38

[210] M. Ye, E. Johns, A. Handa, L. Zhang, P. Pratt, and G.-Z. Yang, "Self-supervised siamese learning on stereo image pairs for depth estimation in robotic surgery," *arXiv preprint arXiv:1705.08260*, 2017. 38

[211] X. Zhang, J. Wang, T. Wang, X. Ji, Y. Shen, Z. Sun, and X. Zhang, "A markerless automatic deformable registration framework for augmented reality navigation of laparoscopy partial nephrectomy," *International Journal of Computer Assisted Radiology and Surgery (IJCARS)*, vol. 14, no. 8, pp. 1285–1294, 2019. 29, 39, 40, 54

[212] Z. Zhang, "Iterative Point Matching for Registration of Free-Form Curves and Surfaces," *International Journal of Computer Vision (IJCV)*, vol. 13, pp. 119–152, 1994. 62, 180, 192

[213] H. Zheng, L. Lin, H. Hu, Q. Zhang, Q. Chen, Y. Iwamoto, X. Han, Y.-W. Chen, R. Tong, and J. Wu, "Semi-supervised segmentation of liver using adversarial learning with deep atlas prior," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Shenzhen, China, 2019, pp. 148–156. 38

[214] H. Zhou and J. Jagadeesan, "Real-time surface deformation recovery from stereo videos," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Shenzhen, China, 2019, pp. 339–347. 21, 29

[215] ——, "Real-time dense reconstruction of tissue surface from stereo optical video," *IEEE Transactions on Medical Imaging (TMI)*, early access, 2019. 15, 24, 25, 54

[216] X.-Y. Zhou, Z.-Y. Wang, P. Li, J.-Q. Zheng, and G.-Z. Yang, "One-stage shape instantiation from a single 2d image to 3d point cloud," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Shenzhen, China, 2019, pp. 30–38. 38, 72, 73

[217] A. Zia, A. Hung, I. Essa, and A. Jarc, "Surgical activity recognition in robot-assisted radical prostatectomy using deep learning," in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, Granada, Spain, 2018, pp. 273–280. 38

[218] M. J. Zinser, H. F. Sailer, L. Ritter, B. Braumann, M. Maegele, and J. E. Zöller, "A Paradigm Shift in Orthognathic Surgery? A Comparison of Navigation, Computer-Aided Designed/Computer-Aided Manufactured Splints, and "Classic" Intermaxillary Splints to Surgical Transfer of Virtual Orthognathic Planning," *Journal of Oral Maxillofacial Surgery (JOMS)*, vol. 71, no. 12, pp. 2151.e1–2151.e21, 2013. 3

Appendix A

Implementation details

A.1 Initialization

During manual initialization the final step is to refine the registration of the surface reconstruction of the initial frame to the preoperative *CT* via the use of a traditional *ICP* technique [15, 29, 212]. The refined registration candidate is viewed again to ensure the physical meaningfulness of the registration has been preserved. Descriptions of the *ICP* parameters and corresponding values that are used across the laboratory and intraoperative datasets and are shown in Table A.1.

The *GO-ICP* semi-automated registration technique has several free parameters. The *Mean Squared Error (MSE)* convergence threshold acts as a termination criteria. If the average squared Euclidean error for all points to the surface falls below the convergence threshold then the algorithm terminates and returns the current transformation. The trim fraction is used when the reconstruction does not perfectly overlap the model (and was disabled for all reported results). Finally, the remaining parameters setup the rotation and translation space evaluated by the

algorithm. For the semi-automated initialization tests, the 50-best candidates were generated and inspected for each of the tested dataset conditions. Descriptions of the *GO-ICP* parameters and values that were used during evaluation are shown in Table A.2. Dataset specific values for the free parameters of the *GO-ICP* module that were used during evaluation are shown in Table A.3.

Variable	Variable description	Value
maximum correspondence distance	Sets the inlier/outlier threshold, i.e., distance > maximum correspondence distance is an outlier	1 mm
maximum iterations	Maximum number of iterations that will be used to try and converge on a solution	300
transformation epsilon	If epsilon between successive transformations < transformation epsilon then return current transformation	1e-8
Euclidean fitness epsilon	If total Euclidean squared error < Euclidean fitness epsilon then terminate	1 mm

Table A.1: Implementation Details - (Initialization) Iterative Closest Point (ICP).

Variable	Variable Description	Value (Range)
MSE convergence threshold	If average squared Euclidean error < *MSE* convergence threshold then the current transformation is returned and the algorithm terminates	0.0001 to 0.0005 mm
rotation minimum	Minimum rotation value in the search space for the algorithm	-3.1416 rad
rotation width	Size of the rotation search space	6.2832 rad
translation minimum	Minimum translation value in the search space for the algorithm	-0.8 to -0.6 mm
translation width	Size of the translation search space	1.2 to 1.6 mm
trim fraction	Percentage of points allowed to be removed from calculation (for use when point cloud only partially overlaps model)	0 (no trimming)
nodes	Nodes per dimension of distance transform cube	300

Table A.2: Implementation Details - Globally Optimal Iterative Closest Point (GO-ICP).

Dataset	*MSE* Convergence Threshold	Rotation		Translation		Trim Fraction	Nodes
		Minimum	Width	Minimum	Width		
laboratory	0.0001 mm	-3.1416 rad	6.2832 rad	-0.6 mm	1.2 mm	0 (no trimming)	300
intraoperative	0.0005 mm			-0.8 mm	1.6 mm		

Table A.3: Implementation Details (By Dataset) - Globally Optimal Iterative Closest Point (GO-ICP).

A.2 Stereo correspondence and 3D surface reconstruction

The stereo correspondence component of the system pipeline contains many of the free parameters most often tuned for a given dataset. A description of the free parameters in the stereo algorithm as well as ranges for how they are set is shown in Table A.4. The baseline shift parameter is part of the preprocessing of the data. The baseline shift acts as a horizontal shift of the rectified, non-reference (left) image in the stereo pair. The shift effectively removes the shift in projection due to the standoff distance of the camera to the patient from the search range. Another variable that is modified across datasets is the window size used for local correspondence computations. Finally, the number of smoothing iterations (using anisotropic diffusion) performed on the resulting disparity map was varied across datasets. Parameters are often tuned in response to the standoff distance between the camera and the surgical site and/or the amount of visible texture on the surface of the liver. The stereo component uses relatively common parameters across the evaluated datasets, which are shown in Table A.5.

The statistical outlier removal component deals with finding points that differ too much from their surroundings and removes them from future processing. A description of the free parameters in this portion of the system as well as typical values are found in Table A.6. For the statistical outlier removal component of the pipeline, standard parameters were used across datasets (with the exception of ***Intraoperative Case 1***). Two parameters were needed for this process: the number of nearest-neighbours to analyze at each point and the outlier threshold (specified as the multiple of the standard deviation of the mean distance to the

query point). The values of the free parameters in the statistical outlier removal component used during evaluation are shown in Table A.7.

For *MLS* smoothing of the data only one free parameter existed. Instead of computing the surface whereby all points in the point cloud were used in a weighted computation for the surface at each point, only points in a nearest-neighbour search range held any influence. All datasets used a search range of 2 mm.

Variable	Variable Description	Value (Range)
baseline shift	The baseline shift acts as a horizontal shift of the rectified, non-reference image in the stereo pair and acts to remove the standoff distance of the camera to the patient from the search range	0 to 300 px
window size	The size of the aggregation window used to determine correspondence	11 to 15 px
search range	The horizontal search range in pixels at each level of the pyramid during *CTF* processing	5 px
pyramid levels	The number of pyramid levels used in the *CTF* approach	5
pyramid type	Prefiltering approach used to prevent aliasing when generating the image pyramids	Laplacian
smoothing iterations	Number of iterations of a boundary preserving disparity smoothing operator via anisotropic diffusion	0 to 10

Table A.4: Implementation Details - Stereo.

Dataset	Baseline Shift	Window Size	Search Range	Pyramid Levels	Pyramid Type	Smoothing Iterations
Laboratory (all)	0 px	15 px	5 px	5	Laplacian	no smoothing
Intraoperative Case 1	150 px					
Intraoperative Case 2	300 px	11 px				10
Intraoperative Case 3						

Table A.5: Implementation Details (By Dataset) - Stereo.

Variable	Variable Description	Value (Range)
number neighbours	The number of nearest neighbours to analyze at each point	500 to 1000 points
outlier threshold	The threshold for inliers/outliers (expressed as the multiple of the standard deviation of the mean distance to the query point)	0.0005

Table A.6: Implementation Details - Statistical Outlier Removal.

Dataset	Number Neighbours	Outlier Threshold
Laboratory (all)	1000	0.0005
Intraoperative Case 1	500	
Intraoperative Case 2 & 3	1000	

Table A.7: Implementation Details (By Dataset) - Statistical Outlier Removal.

A.3 2D feature tracking and generation of a 3D deformation field

Feature tracking is performed by first extracting "good features to track". Implementation details explaining the free parameters, including ranges for how the parameters are set, are shown in Table A.8. Dataset specific values for the free parameters are shown in Table A.9. The maximum number of features that can be returned, minimum quality level of accepted features (expressed as the fraction of the minimal eigenvalue), and minimum distance between returned features (used to help space features over the region to be tracked) are all free parameters. The number of features and feature spacing are modified based on the robustness of the visual structure present on the surface of the liver. As the visual texture decreases, more points are needed at this stage as future motion estimation proves to be more robust to detecting outliers over the surface. When visual texture is distinctive, being more selective on the features maintained for downstream processing proved to be beneficial over the evaluated datasets.

Variable	Variable description	Value (range)
Maximum number of features	Maximum number of features to be returned by the algorithm	10000 to 1000000
Quality level	Minimum quality level of accepted/returned features (expressed as the fraction of the minimal eigenvalue in the gradient feature matrix)	0.001
Minimum distance	Minimum distance between returned features (used to help space features over the region to be tracked)	5 to 20 px

Table A.8: Implementation Details - Good Features to Track.

Dataset	Maximum no. features	Quality level	Minimum distance
Laboratory (all)	100000	0.001	20 px
Intraoperative Case 1	10000		
Intraoperative Case 2	100000		5 px
Intraoperative Case 3	1000000		

Table A.9: Implementation Details (By Dataset) - Good Features to Track.

A.4 6DOF motion estimation and intraoperative alignment

Once a 3D deformation field relating a given frame back to the initial frame has been generated a rigid motion estimation is performed on the two sets of keypoints. As this stage provides a least-squares estimate of the rigid motion relating the two point sets, performing the motion estimation on the full deformation field would not be ideal as it would include any outliers present. The method is made more robust by performing the rigid estimation using a *RANSAC* strategy. All evaluated datasets use common parameters for this component of the pipeline. Descriptions and the common values for the free parameters in this stage, used across all evaluated datasets, are shown in Table A.10.

Variable	Variable Description	Value (Range)
iterations	Number of iterations for the *RANSAC* strategy	20
inlier threshold	The post-registration Euclidean distance between paired points that delineates inliers from outliers	0.15 mm
subsampled features	The number of points subsampled from the two point sets and used for rigid motion estimation at each iteration of the *RANSAC* strategy	30

Table A.10: Implementation Details - Rigid Motion Estimation.

Appendix B

Dataset image sequences

The reference (right) image for each of the reported frames in all of the laboratory and intraoperative dataset conditions is depicted in Figures B.1-B.4.

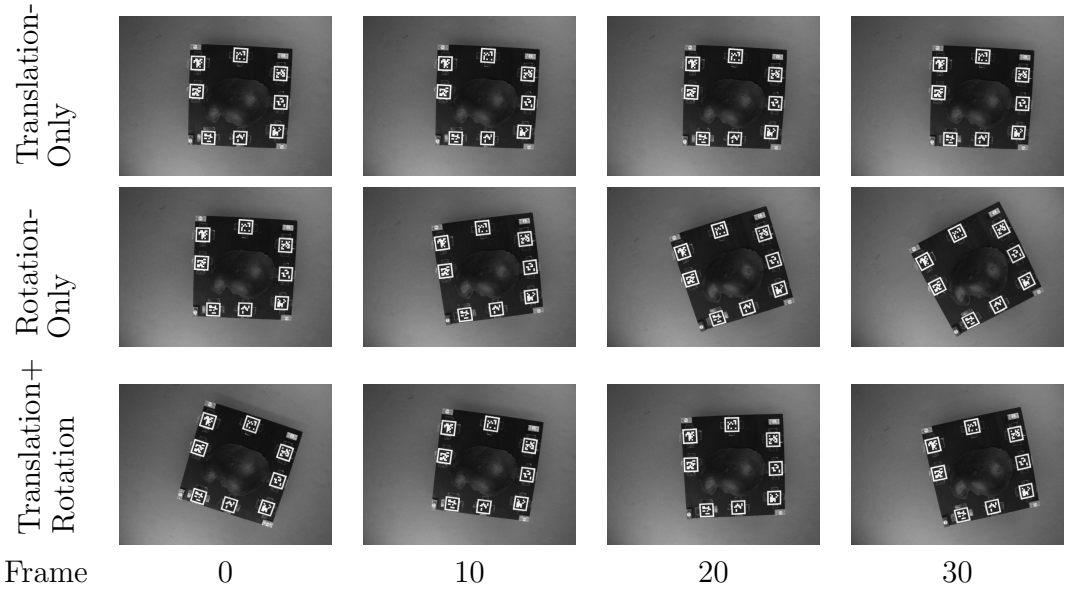


Figure B.1: Laboratory Dataset Image Sequences.

Figure B.2: Intraoperative Dataset Image Sequence: Intraoperative Case 1.

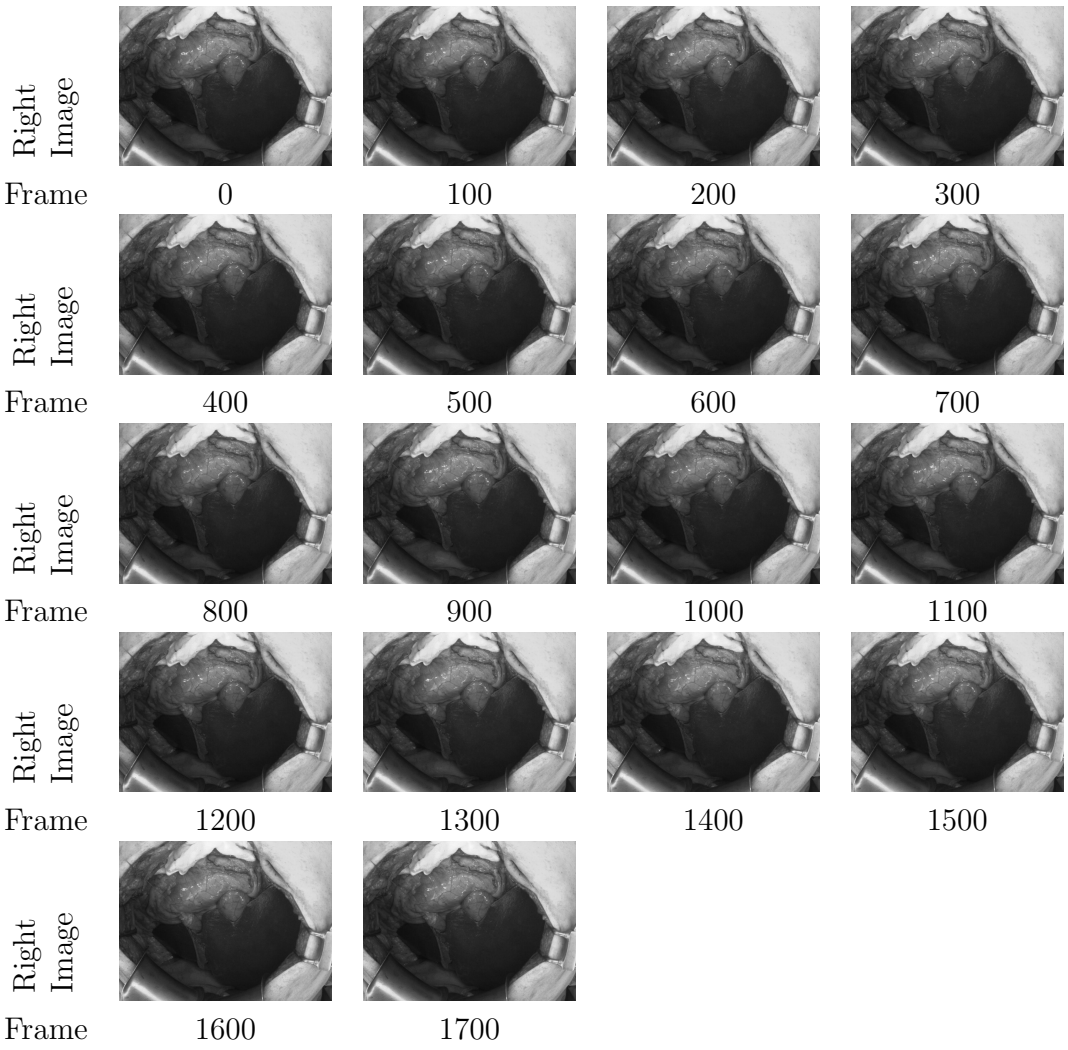


Figure B.3: Intraoperative Dataset Image Sequence: Intraoperative Case 2.

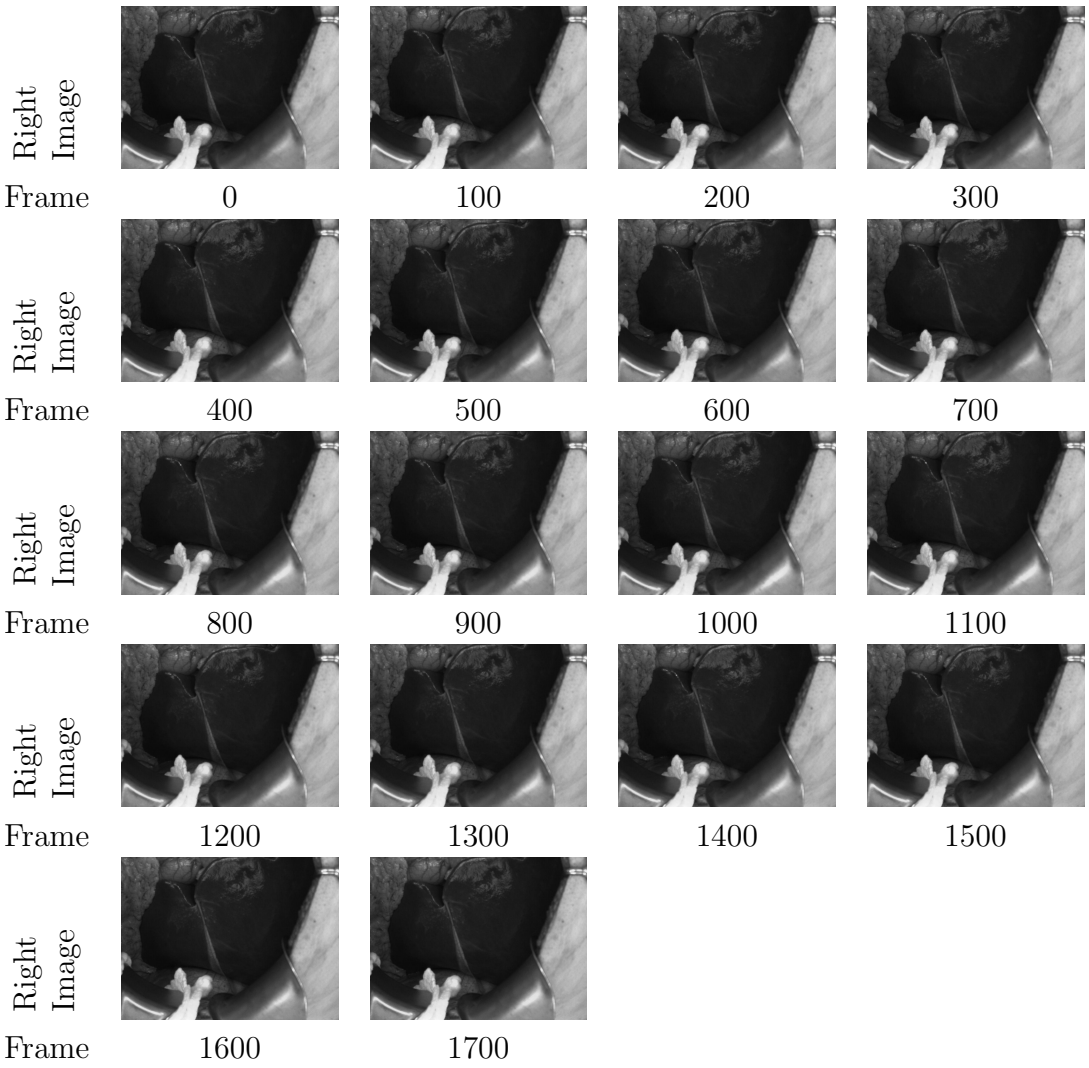


Figure B.4: Intraoperative Dataset Image Sequence: Intraoperative Case 3.

Appendix C

Additional results

Additional tables and figures of results that augment those provided in Chapter 4 are provided here.

SRE statistics on datasets that were previously reported in Section 4.5 using the semi-automated initialization technique are reported here using manual initialization techniques. Registration between *CT* coordinates and the initial frame of each laboratory sequence was performed via identification of five cross-modality features located on the platform supporting the phantom in both the stereo imagery and in the *CT* scan. Intraoperative data was initialized using a manual registration process, specified in Section 3.2.2. Both registration options go through a stage of *ICP* refinement between the *CT* scan of the liver (or phantom liver) and the stereo reconstruction of the visible portion of the *CT* scan to minimize any local error present after initial registration [15, 29, 212]. The resulting registration is reinspected to ensure the alignment is still physically meaningful. ***Intraoperative Case 2*** and ***Intraoperative Case 3*** are not reported in these results as no acceptable semi-automated registration candidate was produced and results using

manual initialization were already presented in Section 4.5. Table C.1 provides average and standard deviation *SRE* statistics using manual initialization. Table C.2 shows the number and percentage of reconstructed points with an *SRE* falling below the thresholds of utility outlined in Section 4.5. The plots in Figures C.1 and C.2 show the distribution of *SRE* over all points in the reconstructed point clouds for the selected frames after manual initialization.

Table C.3 shows the number and the percentage of the reconstructed points that fall within 1 - 5 mm *SRE* ranges for each of the datasets using the preferred initialization method: semi-automated initialization for laboratory datasets and ***Intraoperative Case 1*** and manual registration for ***Intraoperative Case 2*** and ***Intraoperative Case 3***. Similarly to Table 4.5 and related discussion in Chapter 4, instead of only looking at the operationally motivated 5 mm threshold [14], we look at all 1 mm increments below this value (as registration error is only one contribution towards achieving the desired surgical margins). Some of the additional conditions also mark important comparison points. In general, commercial systems rely on optical tracking, which has a stated accuracy of < 2 mm, as evaluated at a single point in time [131]. This bound on the performance of currently approved systems can be used as a point of comparison in Table C.3.

Figure C.3 depicts the front, side, and top views of the registration candidates for ***Intraoperative Case 1*** provided by the modified version of *GO-ICP* presented in Section 3.2.3. The front views are the same as in Figure 4.11. The alternative side and top views reveal degenerate conditions, for example, the surface is registered on the back face of the liver in candidates 3 and 4, which is not clear from the front view alone.

Figures C.4-C.9 show front, side, and top views of the various laboratory and

intraoperative dataset conditions where all reported frames are co-registered in the same visualization. In these figures, each of the frames is given a unique colour (which remains consistent across all of the views). These figures help depict the ability of the presented system to perform interframe registration and the general fit of the intraoperative reconstructions to the preoperative *CT* data.

		Surface Registration Error (SRE) (mm)	
Dataset	Condition	mean	std
Laboratory	Rotation-Only	1.0274	0.8698
	Translation-Only	1.0597	0.8736
	Translation+Rotation	1.0785	0.8998
Intraoperative	Case 1	1.6493	1.4060

Table C.1: Mean and Standard Deviation of Surface Registration Error (SRE) on Datasets with Manual Initialization. Average *SRE* over reported frames are shown with manual initialization of the initial frame reconstruction to the *CT* scan.

Dataset	Condition	Total Points	< 5 mm		< 10 mm		< 15 mm		< 20 mm		< 25 mm	
			# Points	%	# Points	%	# Points	%	# Points	%	# Points	%
Laboratory	Translation-Only	1354374	1350903	99.7	1354374	100	1354374	100	1354374	100	1354374	100
	Rotation-Only	1358820	1356020	99.8	1358820	100	1358820	100	1358820	100	1358820	100
	Translation+Rotation	1344523	1339317	99.6	1344523	100	1344523	100	1344523	100	1344523	100
Intraoperative	Case 1	735806	716043	97.3	734544	99.8	735806	100	735806	100	735806	100

Table C.2: Surface Registration Error (SRE) - Data Within Surgical Tolerance with Manual Initialization. The total number of reconstructed points as well as the number and % of reconstructed points falling within the $<$ 5 - 25 mm *SRE* thresholds are reported here.

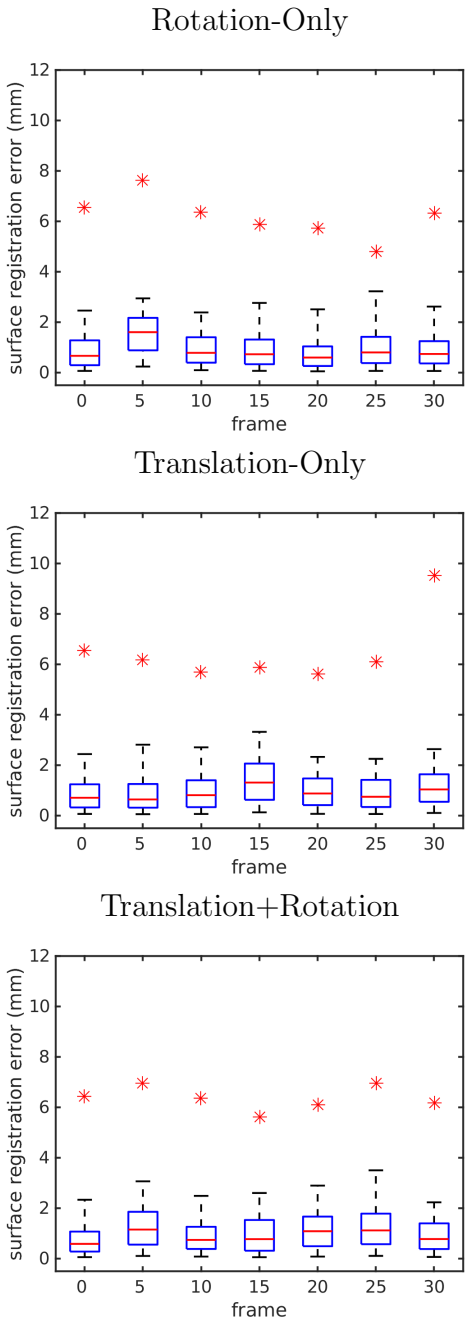


Figure C.1: Surface Registration Error (SRE) - Laboratory Dataset with Manual Initialization. *SRE*, Equation (4.1), is shown at five degree/mm increments across the laboratory sequences. Box plots are shown for each of the reported frames. Blue boxes cover the interquartile range, with the red line showing the median value. Whiskers cover 90% of the reconstructed point clouds. The largest outlier for each dataset is indicated by a red asterisk.

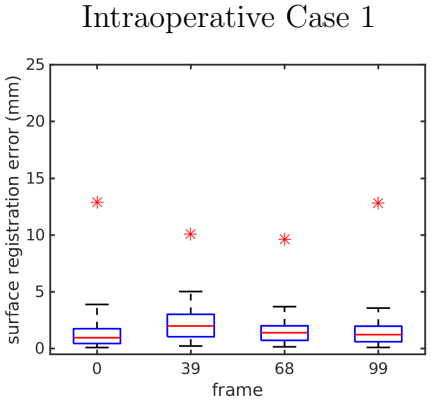


Figure C.2: Surface Registration Error (SRE) - Intraoperative Dataset with Manual Initialization. *SRE*, Equation (4.1), is shown in ***Intraoperative Case 1*** for four frames representing various amounts of respiratory motion that were chosen to show the system performance over the largest portion of the organ's motion. *SRE* is shown at 100 frame increments across the longer intraoperative sequences: ***Intraoperative Case 2*** and ***Intraoperative Case 3***. Box plots are shown for each of the reported frames. Blue boxes cover the interquartile range, with the red line showing the median value. Whiskers cover 90% of the reconstructed point clouds. The largest outlier for each dataset is indicated by a red asterisk.

Dataset	Condition	Total Points	< 1 mm		< 2 mm		< 3 mm		< 4 mm		< 5 mm	
			# Points	%	# Points	%	# Points	%	# Points	%	# Points	%
Laboratory	Translation-Only	1354374	761570	56.2	1174247	86.7	1306724	96.5	1339825	98.9	1350903	99.7
	Rotation-Only	1358820	803918	59.2	1185654	87.3	1311149	96.5	1346416	99.1	1356020	99.8
	Translation+Rotation	1344523	748604	55.7	1159320	86.2	1295681	96.4	1327078	98.7	1339317	99.6
Intraoperative	Case 1	735806	279831	38.0	514558	69.9	636648	86.5	692113	94.1	716043	97.3
	Case 2	2746115	659859	24.0	1193907	43.5	1605754	58.5	1893674	69.0	2120084	77.2
	Case 3	2732357	461229	16.9	1001503	36.7	1541823	56.4	1949909	71.4	2216546	81.1

Table C.3: Surface Registration Error (SRE) - Error Breakdown. The number and percentage of reconstructed points falling within 1 - 5 mm *SRE* are reported across the three laboratory dataset sequences and the three intraoperative dataset sequences. Datasets that had a suitable *GO-ICP* registration candidate (all laboratory cases and ***Intraoperative Case 1***) are initialized using the *GO-ICP* registration candidate evaluated in Section 4.3.2. Otherwise, the manual registration candidate is used for initialization.

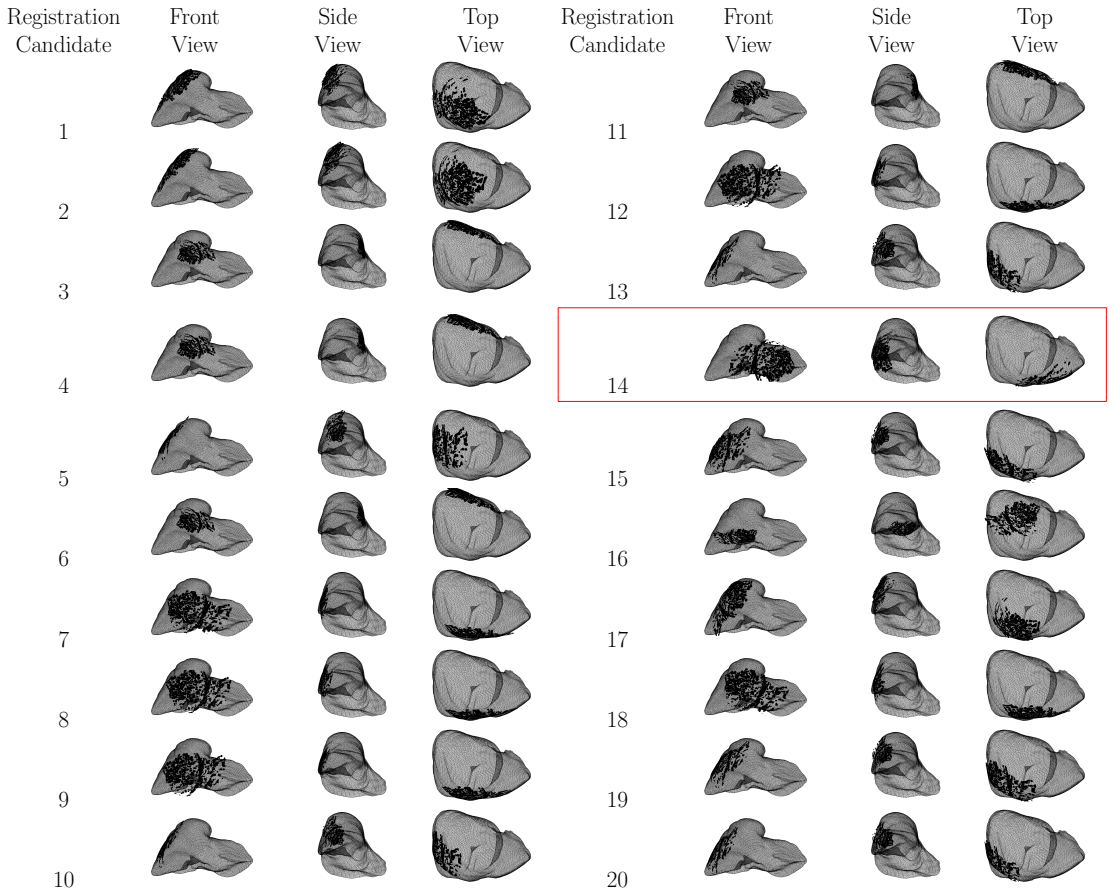


Figure C.3: Globally Optimal Iterative Closest Point (GO-ICP) Registration Initialization Candidates - Intraoperative Case 1.

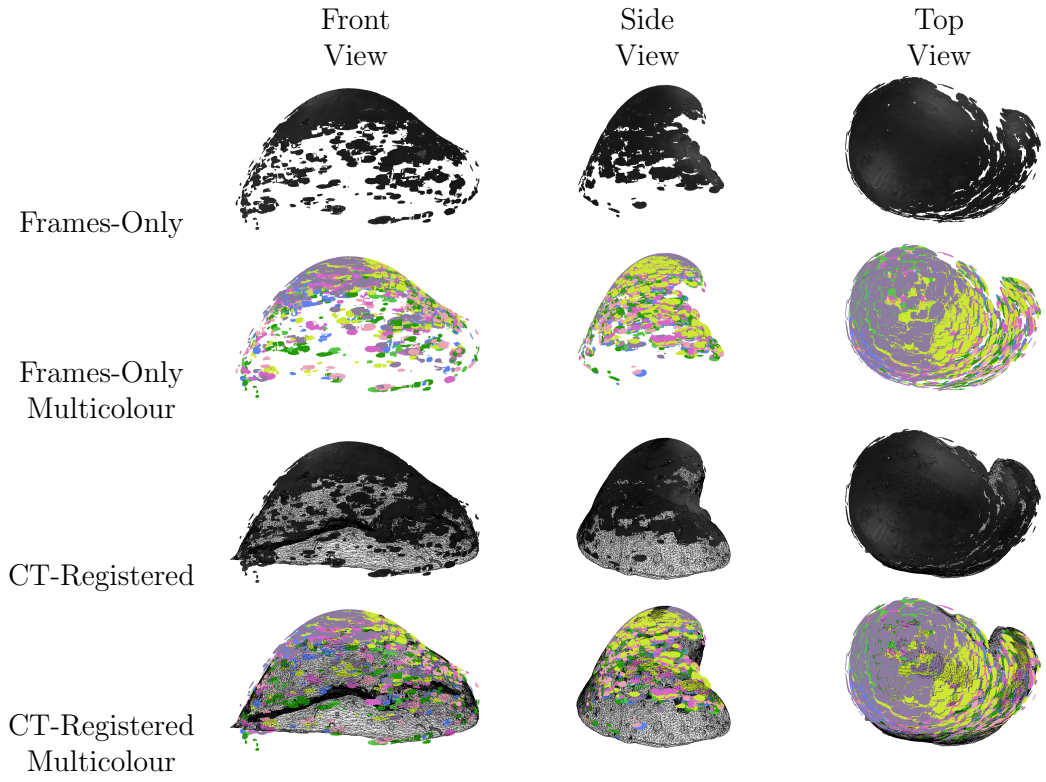


Figure C.4: All Frames Registered - Laboratory Dataset: Translation-Only. The first and third rows show all of the reported frames co-registered, without and with the corresponding *CT* scan included for reference, with the intensity value of each 3D point coming from the 2D intensity at the corresponding pixel in the reference image for that frame. The second and fourth rows show the same 3D information as the first and third rows, respectfully, with the difference being that each reconstructed frame is now assigned a unique colour. This rendering provides a qualitative view of the accuracy of the frame reconstructions and inter-frame registration procedure. Frame colour assignments in the multicolour rows remain consistent across views.

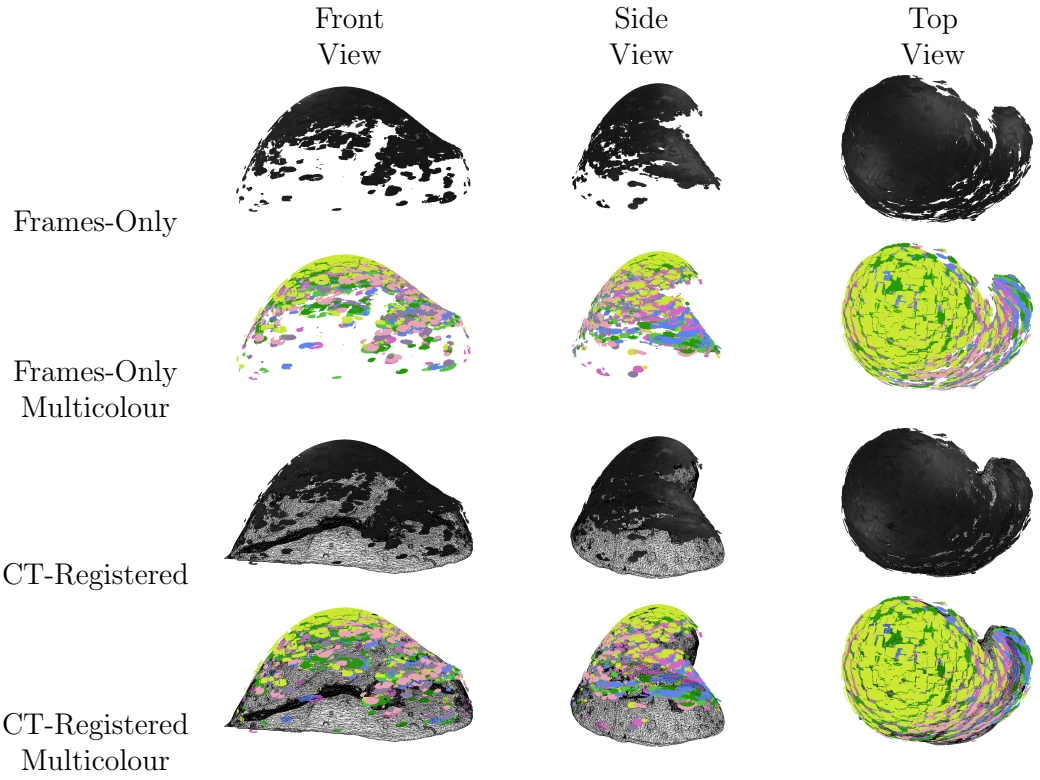


Figure C.5: All Frames Registered - Laboratory Dataset: Rotation-Only. The first and third rows show all of the reported frames co-registered, without and with the corresponding *CT* scan included for reference, with the intensity value of each 3D point coming from the 2D intensity at the corresponding pixel in the reference image for that frame. The second and fourth rows show the same 3D information as the first and third rows, respectfully, with the difference being that each reconstructed frame is now assigned a unique colour. This rendering provides a qualitative view of the accuracy of the frame reconstructions and inter-frame registration procedure. Frame colour assignments in the multicolour rows remain consistent across views.

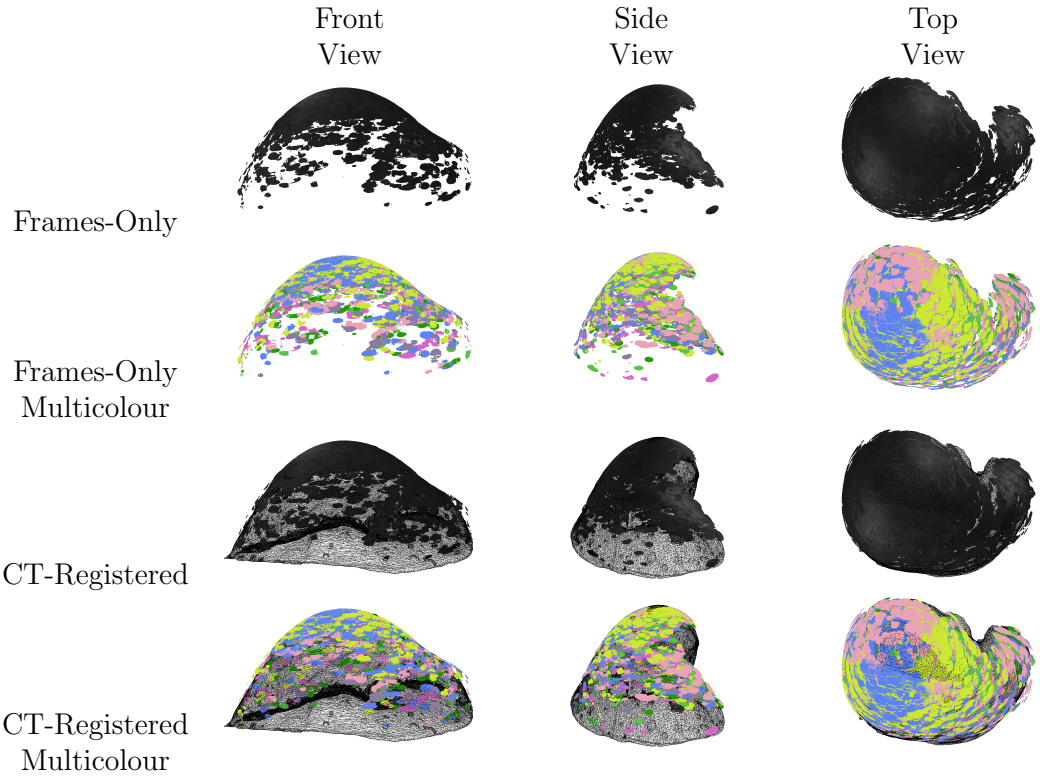


Figure C.6: All Frames Registered - Laboratory Dataset: Translation+Rotation. The first and third rows show all of the reported frames co-registered, without and with the corresponding *CT* scan included for reference, with the intensity value of each 3D point coming from the 2D intensity at the corresponding pixel in the reference image for that frame. The second and fourth rows show the same 3D information as the first and third rows, respectfully, with the difference being that each reconstructed frame is now assigned a unique colour. This rendering provides a qualitative view of the accuracy of the frame reconstructions and inter-frame registration procedure. Frame colour assignments in the multicolour rows remain consistent across views.

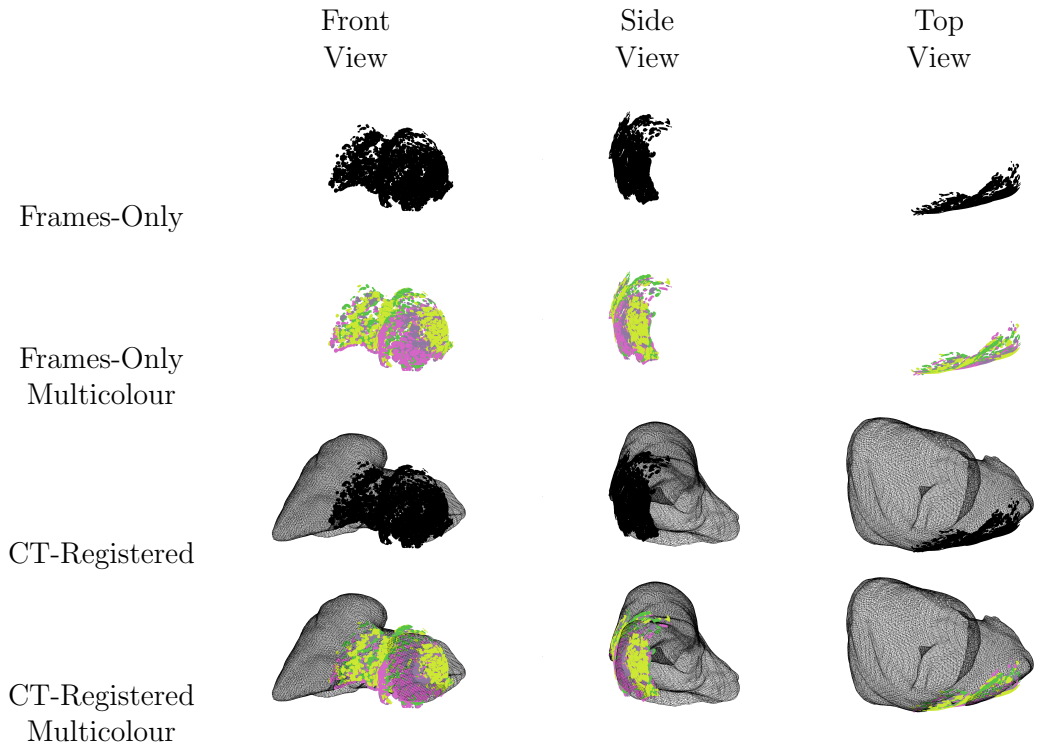


Figure C.7: All Frames Registered - Intraoperative Dataset: Intraoperative Case 1. The first and third rows show all of the reported frames co-registered, without and with the corresponding *CT* scan included for reference, with the intensity value of each 3D point coming from the 2D intensity at the corresponding pixel in the reference image for that frame. The second and fourth rows show the same 3D information as the first and third rows, respectfully, with the difference being that each reconstructed frame is now assigned a unique colour. This rendering provides a qualitative view of the accuracy of the frame reconstructions and inter-frame registration procedure. Frame colour assignments in the multicolour rows remain consistent across views.

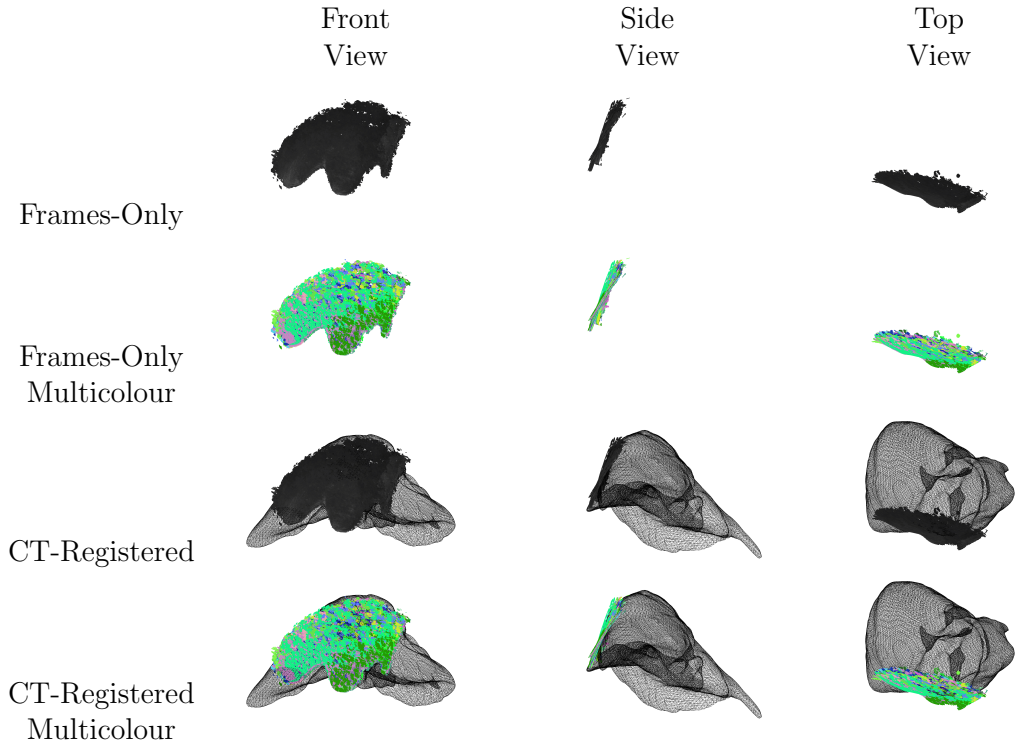


Figure C.8: All Frames Registered - Intraoperative Dataset: Intraoperative Case 2. The first and third rows show all of the reported frames co-registered, without and with the corresponding *CT* scan included for reference, with the intensity value of each 3D point coming from the 2D intensity at the corresponding pixel in the reference image for that frame. The second and fourth rows show the same 3D information as the first and third rows, respectfully, with the difference being that each reconstructed frame is now assigned a unique colour. This rendering provides a qualitative view of the accuracy of the frame reconstructions and inter-frame registration procedure. Frame colour assignments in the multicolour rows remain consistent across views.

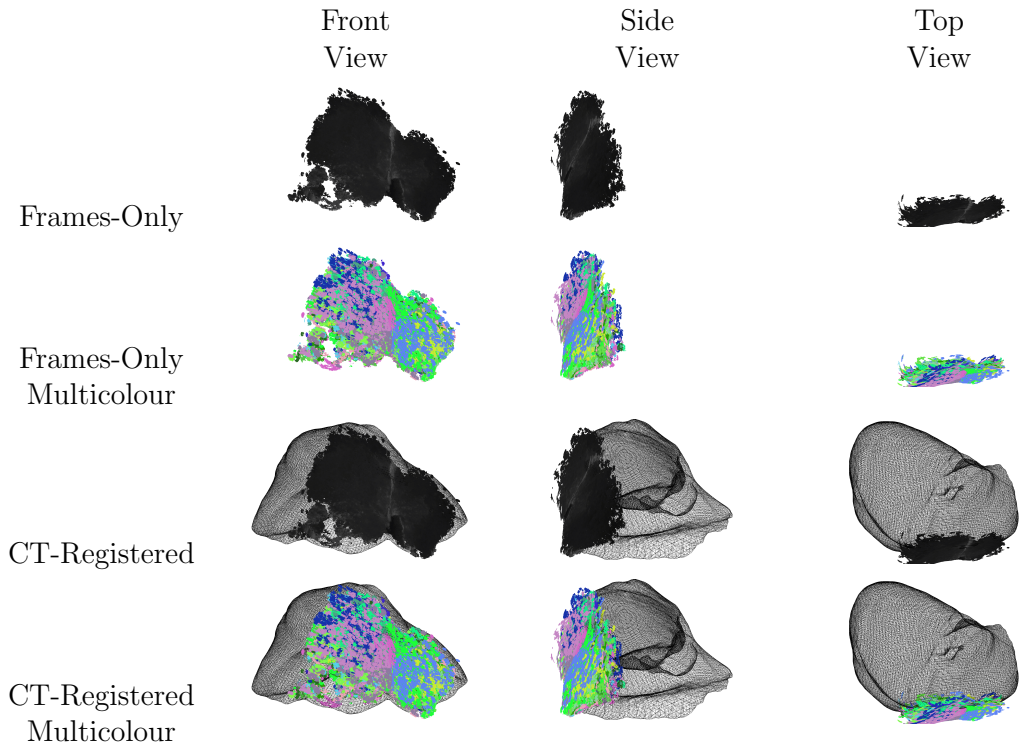


Figure C.9: All Frames Registered - Intraoperative Dataset: Intraoperative Case 3. The first and third rows show all of the reported frames co-registered, without and with the corresponding *CT* scan included for reference, with the intensity value of each 3D point coming from the 2D intensity at the corresponding pixel in the reference image for that frame. The second and fourth rows show the same 3D information as the first and third rows, respectfully, with the difference being that each reconstructed frame is now assigned a unique colour. This rendering provides a qualitative view of the accuracy of the frame reconstructions and inter-frame registration procedure. Frame colour assignments in the multicolour rows remain consistent across views.